PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/47, C12N 15/57, 15/63 — A1

(11) International Publication Number: WO 98/35988

(43) International Publication Date: 20 August 1998 (20.08.98)

(21) International Application Number: PCT/GB98/00415

(22) International Filing Date: 10 February 1998 (10.02.98)

(30) Priority Data:

9703104.1	14 February 1997 (14.02.97)	GB
9722003.2	18 October 1997 (18.10.97)	GB
9722727.6	29 October 1997 (29.10.97)	GB

(71) Applicant *(for all designated States except US)*: ZENECA LIMITED [GB/GB]; 15 Stanhope Gate, London W1Y 6LN (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: EDGE, Michael, Derek [GB/GB]; Zeneca Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG (GB).

(74) Agent: GILES, Allen, Frank; Zeneca Pharmaceuticals, Intellectual Property Dept., Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PROTEINS

(57) Abstract

The field of the invention is recombinant production of carboxypeptidase B. This invention provides a modified prodomain of carboxypeptidase B which enhances recombinant expression thereof when co–expressed from a separate gene. Preferred modified prodomains have added amino acids at their C–terminus, in particular any one of the following sequences: L, KDEL, KKAA or SDYQRL. The carboxypeptidase is preferably human pancreatic carboxypeptidase B. The invention also relates to corresponding polynucleotide sequences, vectors, host cells and methods of recombinant carboxypeptidase B production.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PROTEINS

The present invention relates to modified carboxypeptidase B prodomains and their use in recombinant production of active carboxypeptidase B enzyme.

Many proteases are synthesised *in vivo* as pro-enzymes and the catalytically active protease is eventually cleaved from a larger precursor polypeptide. The pro regions (also called "prodomain" or "pro sequence") of pro-enzymes are usually N-terminal extensions of the mature proteins and in many cases it has been demonstrated that the pro regions are required for folding of their associated protease domains [reviewed by Baker, Shiau and Agard in Current Opinion in Cell Biology (1993) 5, 966-970].

CPB (carboxypeptidase B) has an N-terminal prodomain which is believed to assist correct folding of protein before the prodomain is removed to release active enzyme. Folding of the protein (proenzyme) occurs in the secretory pathway and the prodomain is removed after the proenzyme is secreted from the cell in which it was synthesised. The activation of porcine pancreatic CPB has been studied by V Villegas *et al* in Protein Science (1995), 4, 1792-1800. Activation of procarboxypeptidases has been reviewed by F X Aviles *et al* in Eur. J. Biochem. (1993), 211, 381-389. Folding of recombinantly produced proCPB followed by enzymatic cleavage to produce active CPB has been described in International patent application WO 96/ 23064, Bio-Technology General. CPB is used commercially for example in insulin production and in protein sequencing.

If proCPB is fused at its C terminus to the N-terminus of an antibody chain this allows removal of prodomain (e.g. by trypsin treatment) from the N-terminus of the fusion construct. Alternatively if proCPB was attached to the C-terminus of an antibody chain then the problem arises of how to remove the prodomain from the "middle" of the construct. The solution is to co-express the prodomain separately (*in trans*). Accordingly we have shown that by independently expressing the pro and protease domains of CPB, active human CPB (HCPB) is secreted from COS cells (see International Patent Application WO 96/20011). Moreover, when LC/Fd-HCPB antibody fusion proteins are expressed from COS cells along with independent expression of the prodomain, fully functional Fab'-HCPB and $F(ab')_2$-$(HCPB)_2$ proteins are secreted. Activation of a protease by independent expression of its prodomain has also been reported for a serine protease from *Lysobacter enzymogenes* by Silen

et al in Nature (1989), 341, 462 but this same approach failed in the case of rat pancreatic carboxypeptidase A1 (Phillips *et al*, Biochemistry (1996), 35, 6771-6776. International Patent Application WO 97/42329, Zeneca Limited, published 13-Nov-97 (*i.e.* after the priority dates of the present invention), uses a modified CPB prodomain in Example 48(c); however the art to date has been otherwise totally silent regarding modification of the CPB prodomain.

The present invention is based on the discovery that modification of the prodomain sequence leads to improved recombinant expression yields of active carboxypeptidase B.

According to one aspect of the present invention there is provided a modified prodomain of carboxypeptidase B which enhances recombinant expression thereof when co-expressed from a separate gene.

The term "prodomain" means an N-terminal sequence which is naturally responsible for assisting the folding of CPB into its active conformation before being subsequently removed to give active mature CPB. The natural prodomain of pancreatic HCPB is set out in SEQ ID NO: 12 at positions 14-108. The sequence of mature pancreatic HCPB is set out in SEQ ID NO: 12 at positions 109-415. The prodomain attached to mature pancreatic HCPB is termed "proHCPB" and is set out in SEQ ID NO: 12 at positions 14-415.

The term "modified prodomain" in this context means that the prodomain is different from the naturally occurring prodomain for example by addition, deletion, substitution or insertion of amino acids. An example of a suitable substitution is to change the Arg residue at the C-terminus of the prodomain of pancreatic HCPB (at position 108 in in SEQ ID NO: 12) into a hydrophobic amino acid such as leucine. Suitable C-terminus additions are described herein for HCPB. Suitable deletions include deletion of any of the last 3 amino acids of the natural prodomain of HCPB (positions 106-108 in SEQ ID NO: 12). A skilled worker in the field will be able to select and test further modified prodomains based on the disclosures herein with reference in particular to the specific examples and proposed mechanism of the effect.

In this specification all amino acid sequences are preferably of L-configuration.

The term "enhances recombinant expression" means that modifications tested by measuring expression levels of carboxypeptidase B in the presence of a modified prodomain compared with unmodified (i.e. natural) prodomain; enhanced levels of expression of at least

20 %, more preferably at least 50 % are within the scope of the invention. Suitable expression systems for evaluation of enhanced expression include those set out in Reference Example 4 (for CPB) and Examples 14 & 19 (for "reversed polarity" mutants of CPB).

Preferably the prodomain is modified at its C-terminus, more preferably by addition of amino acids, more preferably by addition of at least one amino acid, more preferably by addition of 1-20 amino acids, more preferably by addition of 1-15 amino acids, more preferably by addition of 1-10 amino acids and especially by addition of 1-6 amino acids. Preferably the C-terminus amino acid of the prodomain after addition of the amino acid(s) is a hydrophobic amino acid. Preferably the C-terminus of the added amino acids is a hydrophobic amino acid. Preferred hydrophobic amino acids are leucine, isoleucine, valine, alanine or phenylalanine and especially leucine. Preferred specific sequences are addition of any one of L, KDEL. KKAA or SDYQRL sequences and of these KKAA, SDYQRL or L are preferred, with L being especially preferred.

CPB has been reviewed in The Worthington Manual (1988), pages 65-67, Worthington Biochemical Corporation, Freehold, New Jersey, USA 07728 and by J E Folk in The Enzymes, Volume III, 1971, pages 57-79, Academic Press, New York (Ed. P D Boyer). Preferably the carboxypeptidase is human and especially human pancreatic CPB. Thus a preferred modified prodomain is a human pancreatic carboxypeptidase B prodomain with a leucine added at its C-terminus.

The carboxypeptidase enzyme may be fused to another protein such as for example an antibody chain. Another suitable carboxypeptidase is plasma carboxypeptidase B (Eaton, D.L. et al., J.Biol. Chem. 1991 266 21833-38). CPB generally catalyses hydrolysis of basic amino acids from the C-terminus of polypeptides. Carboxypeptidase mutants are also within the scope of the invention. Carboxypeptidase mutants include enzymes having altered substrate specificity. In International Patent Application WO 96/20011, published 4-Jul-96, we proposed a "reversed polarity" ADEPT system based on mutant human enzymes having the advantage of low immunogenicity compared with for example bacterial enzymes. A particular host enzyme was human pancreatic CPB (see for example, Example 15 [D253K]human CPB & 16 [D253R]human CPB therein) and prodrugs therefor (see Examples 18 & 19 therein). The host enzyme is mutated to give a change in mode of interaction between enzyme and prodrug in terms of recognition of substrate compared with

the native host enzyme. In our subsequent International Patent Application WO 97/07769 (published 6-Mar-97) further work on mutant CPB enzyme/prodrug combinations for ADEPT are described. Preferred "reversed polarity" mutant carboxypeptidases are any one of [D253K]HCPB, [G251T,D253K]HCPB or [A248S,G251T,D253K]HCPB of which the latter is most preferred. The carboxypeptidase enzyme may be in the form of a recombinant fusion protein with another peptide such as for example an antibody heavy or light chain.

In another embodiment of the invention the carboxypeptidase B is porcine. Expression of porcine procarboxypeptidase B has been described in International patent application WO 95/14096, Eli Lilly.

Without wishing to be bound by theoretical considerations the following text sets out our understanding of the mechanism behind the invention. The C-terminal residue of the natural prodomain of pancreatic CPB is arginine. This is known to be removed readily by active CPB itself or by other enzymes for example carboxypeptidase H which is believed to be present in the Golgi apparatus (which along with the endoplasmic reticulum are the major organelles involved in intracellular trafficking of secreted proteins). This degradation is believed to reduce any interaction between the prodomain and mature enzyme. Thus we believe that if the degradation can be prevented, folding and secretion of CPB protein will be enhanced. For example, addition of a C-terminal amino-acid (e.g. leucine) residue not cleaved by the active enzyme is believed to hinder degradation. Another potential mechanism of enhanced expression is based on enhanced intracellular trafficking. Protein folding in eucaryotic cells takes place in the endoplasmic reticulum so modification of the prodomain with sequences (e.g. KKAA) which increase the concentration of the prodomain in intracellular compartments, especially the endoplasmic reticulum, are also believed to enhance expression yields of recombinantly expressed CPB.

According to another aspect of the present invention there is provided a polynucleotide sequence capable of encoding a modified prodomain of the invention.

According to another aspect of the present invention there is provided a vector comprising a polynucleotide sequence capable of encoding a modified prodomain of the invention.

According to another aspect of the present invention there is provided a host cell comprising a polynucleotide sequence, optionally in the form of a vector, capable of encoding a modified prodomain of the invention.

According to another aspect of the present invention there is provided method of recombinant carboxypeptidase B production which comprises simultaneously expressing in a host cell a carboxypeptidase B enzyme gene together with a separate gene encoding a modified prodomain of the invention and optionally at least partially purifying the recombinant carboxypeptidase B. The carboxypeptidase B enzyme gene can be in the form of mature CPB enzyme or proCPB (with natural prodomain). In the case of the latter, without wishing to be bound by theoretical considerations, it is contemplated that separately expressed modified prodomain is able to interact with expressed CPB through participation in a dynamic equilibrium. The host cell can be procaryotic or eucaryotic, preferably eucaryotic. more preferably mammalian and most especially CHO cells. In another embodiment of the invention an especially preferred host cell is in the form of a transgenic animal. The separate genes may be on the same or a different genetic entity e.g. the same or different plasmids.

The carboxypeptidase B enzyme may be in the form of a recombinant fusion protein with another peptide such as for example an antibody heavy or light chain. An especially preferred carboxypeptidase B fusion construct is described in Example 15 below. Example 15 describes a gene encoding a humanised Fd heavy chain fragment of antibody 806.077 linked to enzyme [A248S,G251T,D253K]HCPB and its co-expression with a gene encoding a humanised light chain of 806.077 and a gene encoding the pro-L modified prodomain of human carboxypeptidase B to give the $F(ab')_2$ protein with a molecule of [A248S,G251T,D253K]HCPB at the C-terminus of each of the heavy chain fragments. The constant and hinge regions of of the humanised Fd heavy chain fragment are derived from the human IgG3 antibody isotype.

Accordingly a preferred method of recombinant carboxypeptidase B production is one in which the eucaryotic host cell is mammalian and:

i) the recombinant carboxypeptidase B is in the form of a humanised 806.077 $F(ab')_2$ - $\{[A248S,G251T,D253K]HCPB\}_2$ fusion protein;

ii) the carboxypeptidase B enzyme gene is in the form of a gene encoding

a humanised Fd heavy chain fragment of antibody 806.077 linked to enzyme [A248S,G251T,D253K]HCPB;

iii) the separate gene encoding a modified prodomain encodes human pancreatic carboxypeptidase B prodomain with a leucine added at its C-terminus; and

iv) a further gene is co-expressed which encodes a humanised light chain of antibody 806.077;

and wherein the fusion protein is in the form of a $F(ab')_2$ with a molecule of [A248S,G251T,D253K]HCPB at a C-terminus of each of its heavy chain fragments.

Hybridoma 806.077 antibody was deposited at the European Collection of Animal Cell Cultures (ECACC), PHLS Centre for Applied Microbiology & Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom on 29th February 1996 under accession no. 96022936 in accordance with the Budapest Treaty. Humanisation of antibody 806.077 has been described in International Patent Application WO 97/42329, Zeneca Limited, published 13-Nov-97.

Some expression systems involve transforming a host cell with a vector; such systems are well known such as for example in E. coli, yeast and mammalian hosts (see Methods in Enzymology 185, Academic Press 1990). Other systems of expression are also contemplated such as for example transgenic non-human mammals in which the gene of interest, preferably cut out from a vector and preferably in association with a mammary promoter to direct expressed protein into the animal's milk, is introduced into the pronucleus of a mammalian zygote (usually by microinjection into one of the two nuclei (usually the male nucleus) in the pronucleus) and thereafter implanted into a foster mother. A proportion of the animals produced by the foster mother will carry and express the introduced gene which has integrated into a chromosome. Usually the integrated gene is passed on to offspring by conventional breeding thus allowing ready expansion of stock. Preferably the protein of interest is simply harvested from the milk of female transgenic animals. The reader is directed to the following publications: Simons et al. (1988), Bio/Technology 6:179-183; Wright et al. (1991) Bio/Technology 9:830-834; US 4,873,191; WO 95/17085 (Genzyme Transgenics) and; US 5,322,775. Manipulation of mouse embryos is described in Hogan et al. "Manipulating the Mouse Embryo; A Laboratory Manual", Cold Spring Harbor Laboratory 1986. Further references on transgenic animal production techniques include the following:

(1) Center for Biologics Evaluation and Research, "Points to Consider in the Manufacture and Testing of Therapeutic Products for Human Use Derived from Transgenic Animals" (Food and Drug Administration, Rockville, MD, 1995).

(2) "Use of Transgenic Animals in the Manufacture of Biological Medicinal Products for Human Use", Ad hoc Working Party on Biotechnology/Pharmacy, Directorate-General III/3612/93 Final (1995).

(3) L.-M. Houdebine, "Production of Pharmaceutical Proteins from Transgenic Animals," *J. Biotechnol.* 34, 269-287 (1994).

(4) Y. Echelard, "Recombinant Protein Production in Transgenic Animals," *Curr. Op. Biotechnol.* 7, 536-540 (1996).

(5) E.A. Maga and J.D. Murray, "Mammary Gland Expression of Transgenes and the Potential for Altering the Properties of Milk," *Bio/Technology* 13, 1452-1457 (1995).

(6) C. Ziomek, "Minimization of Viral Contamination in Human Pharmaceuticals Produced in the Milk of Transgenic Goats," *Dev. Biol. Stand.* 88, 263-266 (1996).

(7) S. Groet and H. Meade, "Antithrombin III: Clinical Development Results and Future Plans," *Abstract, IBC Symposium: Transgenic Therapeutics*, West Palm Beach. FL, 5-6 February 1997.

(8) M. Hayes et al., "Recombinant Therapeutic Protein Recovery from Transgenic Milk," *Abstract, Recovery of Biological Products VIII*, Tucson, AZ, 20-25 October 1996.

(9) Young et al., "Production of Biopharmaceutical Proteins in Milk of Transgenic Dairy Animals," *Bio Pharm*, Volume 10, Number 6, pages 34-38, February 1997.

Transgenic plant technology is also contemplated such as for example described in the following publications: Swain W.F. (1991) TIBTECH 9: 107-109; Ma J.K.C. et al (1994) Eur. J. Immunology 24: 131-138; Hiatt A. et al (1992) FEBS Letters 307:71-75; Hein M.B. et al (1991) Biotechnology Progress 7: 455-461; Duering K. (1990) Plant Molecular Biology 15: 281-294.

If desired, host genes can be inactivated or modified using standard procedures as outlined briefly below and as described for example in "Gene Targeting; A Practical Approach", IRL Press 1993. The target gene or portion of it is preferably cloned into a vector with a selection marker (such as Neo) inserted into the gene to disrupt its function. The vector is linearised then transformed (usually by electroporation) into embryonic stem (ES) cells (eg

derived from a 129/Ola strain of mouse) and thereafter homologous recombination events take place in a proportion of the stem cells. The stem cells containing the gene disruption are expanded and injected into a blastocyst (such as for example from a C57BL/6J mouse) and implanted into a foster mother for development. Chimaeric offspring can be identified by coat colour markers. Chimeras are bred to ascertain the contribution of the ES cells to the germ line by mating to mice with genetic markers which allow a distinction to be made between ES derived and host blastocyst derived gametes. Half of the ES cell derived gametes will carry the gene modification. Offspring are screened (eg by Southern blotting) to identify those with a gene disruption (about 50 % of progeny). These selected offspring will be heterozygous and therefore can be bred with another heterozygote and homozygous offspring selected thereafter (about 25 % of progeny). Transgenic animals with a gene knockout can be crossed with transgenic animals produced by known techniques such as microinjection of DNA into pronuclei, sphaeroplast fusion (Jakobovits et al. (1993) Nature 362:255-258) or lipid mediated transfection (Lamb et al. (1993) Nature Genetics 5 22-29) of ES cells to yield transgenic animals with an endogenous gene knockout and foreign gene replacement.

ES cells containing a targeted gene disruption can be further modified by transforming with the target gene sequence containing a specific alteration, which is preferably cloned into a vector and linearised prior to transformation. Following homologous recombination the altered gene is introduced into the genome. These embryonic stem cells can subsequently be used to create transgenics as described above.

The term "host cell" includes any procaryotic or eucaryotic cell suitable for expression technology such as for example bacteria, yeasts, plant cells and non-human mammalian zygotes, oocytes, blastocysts, embryonic stem cells and any other suitable cells for transgenic technology. If the context so permits the term "host cell" also includes a transgenic plant or non-human mammal developed from transformed non-human mammalian zygotes, oocytes, blastocysts, embryonic stem cells, plant cells and any other suitable cells for transgenic technology.

In this specification conservative amino acid analogues of specific prodomain amino acid sequences are contemplated which substantially retain the properties of a prodomain of the invention but differ in sequence by one or more conservative amino acid substitutions, deletions or additions. It is known in the art that N-terminus additions can often be

accommodated. A skilled worker in the field will be able to select and test conservative analogues using disclosures herein together with general knowledge. For example, the 3 dimensional structure of porcine CPB is known from Coll et al, 1991, The EMBO Journal, 10, 1-9. A skilled worker can use the 3D structure to select potential non-critical regions for analogue production. For example, non-critical loop regions may be selected for analogue production.

However the specifically listed amino acid sequences are preferred. Typical conservative amino acid substitutions are tabulated below.

Conservative Substitutions

Original	Exemplary Substitutions	Preferred Substitutions
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Lys; Arg	Gln
Asp (D)	Glu	Glu
Glu (E)	Asp	Asp
Gly (G)	Pro	Pro
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe; Norleucine	Leu
Leu (L)	Norleucine; Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Leu; Val; Ile; Ala	Leu
Pro (P)	Gly	Gly
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Tyr (Y)	Trp; Phe; Thr; Ser	Phe

Original	Exemplary Substitutions	Preferred Substitutions
Val (V)	Ile; Leu; Met; Phe; Ala; Norleucine	Leu

In this specification nucleic acid variations (deletions, substitutions and additions) of specific nucleic acid sequences are contemplated which retain which the ability hybridise under stringent conditions to the specific sequence in question. However specifically listed nucleic acid sequences are preferred.

Abbreviations used herein include:

ADEPT	antibody directed enzyme prodrug therapy
CPB	carboxypeptidase B
DAB	substrate 3,3'-diaminobenzidine tetrahydrochloride
DEPC	diethylpyrocarbonate
DMEM	Dulbecco's modified Eagle's medium
ECACC	European Collection of Animal Cell Cultures
EIA	enzyme immunoassay
ELISA	enzyme linked immunosorbent assay
FCS	foetal calf serum
Fd	heavy chain of Fab, Fab' or $F(ab')_2$ optionally containing a hinge
HCPB	human carboxypeptidase B, preferably pancreatic
HRPO	horse radish peroxidase
LC	antibody light chain
NCIMB	National Collections of Industrial and Marine Bacteria
PBS	phosphate buffered saline
PCR	polymerase chain reaction
proCPB	CPB with prodomain attached
SDS-PAGE	sodium dodecyl sulphate - polyacrylamide gel electrophoresis
TBS	Tris-buffered Saline

The invention is illustrated by the following non-limiting Examples (supported by Reference Examples) in which:

Figure 1 illustrates pancreatic HCPB cloning.

Figure 2 illustrates pancreatic HCPB sequencing.

Figure 3 illustrates vector pICI1266.

Figure 4 illustrates pICI1266 expression vector gene cloning.

DNA is recovered and purified by use of GENECLEAN™ II kit (Stratech Scientific Ltd. or Bio 101 Inc.). The kit contains: 1) 6M sodium iodide; 2) a concentrated solution of sodium chloride, Tris and EDTA for making a sodium chloride/ethanol/water wash; 3) Glassmilk- a 1.5 ml vial containing 1.25 ml of a suspension of a specially formulated silica matrix in water. This is a technique for DNA purification based on the method of Vogelstein and Gillespie published in Proceedings of the National Academy of Sciences USA (1979) Vol 76, p 615. Briefly, the kit procedure is as follows. To 1 volume of gel slice is added 3 volumes of sodium iodide solution from the kit. The agarose is melted by heating the mix at 55^{o}C for 10 min then Glassmilk (5-10ml) is added, mixed well and left to stand for 10 min at ambient temperature. The glassmilk is spun down and washed 3 times with NEW WASH (0.5ml) from the kit. The wash buffer is removed from the Glassmilk which is to dry in air. The DNA is eluted by incubating the dried Glassmilk with water (5-10ml) at 55^{o}C for 5-10 min. The aqueous supernatant containing the eluted DNA is recovered by centrifugation. The elution step can be repeated and supernatants pooled;

Competent E. coli DH5a cells were obtained from Life Technologies Ltd (MAX efficiency DH5α competent cells);

Serum free medium is OPTIMEM™ I Reduced Serum Medium, GibcoBRL Cat. No. 31985;

Mini-preparations of double stranded plasmid DNA were made using the RPM™ DNA preparation kit from Bio101 Inc. (cat. No 2070-400) or a similar product - the kit contains alkaline lysis solution to liberate plasmid DNA from bacterial cells and glassmilk in a spinfilter to adsorb liberated DNA which is then eluted with sterile water or 10 mM Tris-HCl, 1 mM EDTA, pH 7.5;

LIPOFECTIN™ Reagent (GibcoBRL Cat. No. 18292-011) is a 1:1 (w/w) liposome formulation of the cationic lipid N-[1-(2,3-dioleyloxy)propyl]-n,n,n-trimethylammonium

chloride (DOTMA) and dioleoyl phosphatidylethanolamine (DOPE) in membrane filtered water. It binds spontaneously with DNA to form a lipid-DNA complex - see Felgner et al. in Proc. Natl. Acad. Sci. USA (1987) 84, 7431;

Oligonucleotide sequences were prepared in an Applied Biosystems DNA synthesiser from 5'dimethoxytrityl base-protected nucleoside-2-cyanoethyl-N,N'-di-isopropyl-phosphoramidites and protected nucleoside linked to controlled-pore glass supports on a 0.2 µmol scale, according to the protocols supplied by Applied Biosystems Inc.;

AMPLITAQ™ ,available from Perkin-Elmer Cetus, is used as the source of thermostable DNA polymerase;

T4 DNA ligase was obtained from New England Biolabs Inc.; and

General molecular biology procedures can be followed from any of the methods described in "Molecular Cloning - A Laboratory Manual" Second Edition, Sambrook. Fritsch and Maniatis (Cold Spring Harbor Laboratory, 1989).

Laemmli Loading Buffer is 0.125 M Tris-HCl pH 6.8, containing 2 % SDS, 2 % β -mercaptoethanol, 10 % glycerol and 0.1 % Bromophenol blue.

Buffer A is 200 mM Tris (hydroxymethyl)aminomethane hydrochloride (TRIS-HCl), 20 % sucrose, pH 8.0.

Elution Buffer is 100 mM sodium carbonate, 500 mM sodium chloride, pH 11.4.

Lysis Buffer is 50 mM Tris-HCl, 15 % sucrose, pH 8.0.

10X Enzyme Buffer is 500 mM KCl, 100 mM Tris (pH8.3), 15 mM $MgCl_2$ and 0.1 % gelatin.

PCR Buffer is 10 mM Tris-HCl (pH 8.3), 50 mM KCL, 1.5 mM $MgCl_2$, 0.125 mM each of dATP, dCTP, dGTP and dTTP.

Phosphate Buffer is 50 mM phosphate buffer, pH 6.5.

All temperatures are in degrees centigrade.

Reference Example 1

Cloning of human pancreatic carboxypeptidase B (HCPB)

The coding sequence for human pancreatic carboxypeptidase B was obtained from a human pancreatic cDNA library cloned in the λgt10 vector (Clontech, Human pancreas

5' STRETCH cDNA, HL1163a) using PCR technology, and cloned into the plasmid vector pBluescript II KS+ (Stratagene).

Typically, an aliquot of the cDNA library (5µl at a titre of $>10^8$pfu/ml) was mixed with 100pMols of two oligonucleotide primers, BPT1 and BPB1, (SEQ ID NO: 1 and SEQ ID NO: 2), dNTPs to a final concentration of 200µM, thermostable DNA polymerase reaction buffer, and 2.5U of thermostable DNA polymerase in a final volume of 100µl. The mixture was heated at 94° for 10 minutes prior to addition to the thermostable DNA enzyme, and the PCR incubation was carried out using 30 cycles of 94° for 1.5 minutes, 50° for 2 minutes, and 72° for 2 minutes, followed by a single incubation of 72° for 9.9 minutes at the end of the reaction.

The two oligonucleotide primers were designed to allow PCR extension from the 5' of the gene from BPT1 (SEQ ID NO: 1), between the start of the pre-sequence and the start of the pro-sequence, and PCR extension back from the 3' end of the gene from BPB1(SEQ ID NO: 2), as shown in **Figure 1**. BPT1 and BPB1 are also designed to introduce unique restriction sites, SacI and XhoI respectively, into the PCR product.

An aliquot of the PCR product was analysed for DNA of the correct size (about 1250 base pairs) by agarose gel electrophoresis and found to contain predominantly a band of the correct size. The remainder of the product from the reaction mix was purified and separated from excess reagents using a microconcentrator column (Centricon 100, Amicon), followed by DNA isolation by ethanol/sodium acetate precipitation, centrifugation, vacuum drying and re-suspension in distilled water. The isolated DNA was restriction digested with enzymes SacI and XhoI, and a band of the correct size (about 1250 base pairs) purified and isolated.

pBluescript II KS+ double stranded DNA (Stratagene) was restriction digested with SacI enzyme, and the product dephosphorylation treated with calf intestinal alkaline phosphatase to remove 5'phosphoryl groups and reduce re-ligation and vector background following transformation. The DNA product was purified from enzyme reaction contaminants using glass-milk, and then restriction digested with XhoI enzyme. DNA of the correct size (about 2850 base pairs) was purified.

Aliquots of both restricted and purified DNA samples were checked for purity and concentration estimation using agarose gel electrophoresis compared with known

standards. From these estimates ligation mixes were prepared to clone the HCPB gene into the vector, using a molar ratio of about 1 vector to 2.5 insert (1 pBluescript II KS+ to 2.5 HCPB PCR product), and a final DNA concentration of about 2.5 ng/µl, in the presence of T4 DNA ligase, 1 mM ATP and enzyme buffer.

Following the ligation reaction the DNA mixture was used to transform E.coli strain DH5α. Cell aliquots were plated on L-agar nutrient media containing 100 µg/ml ampicillin as selection for plasmid vector, and incubated over-night at 37°. Colonies containing plasmids with inserts of interest were identified by hybridisation.

About 200 colonies were picked and plated onto duplicate sterile nitro-cellulose filters (Schleicher and Schull), pre-wet on plates of L-agar nutrient media containing 100µg/ml ampicillin as selection for plasmid vector, and incubated over-night at 37°. One duplicate plate was stored at 4°, and acted as a source of live cells for the colonies, the other plate was treated to denature and fix the DNA from the individual colonies to the nitro-cellulose. The nitro-cellulose filter was removed from the agar plate and placed in succession onto filter papers (Whatman) soaked in:

1. 10 % SDS for 2 minutes;
2. 0.5M NaOH, 1.5M NaCl for 7 minutes
3. 0.5M NaOH, 1.5M NaCl for 4 minutes
4. 0.5M NaOH, 1.5M NaCl for 2 minutes
5. 0.5M Tris pH7.4, 1.5M NaCl for 2 minutes
6. 2xSSC (standard saline citrate) for 2 minutes.

The filter was then placed on a filter paper (Whatman) soaked in 10xSSC and the denatured DNA cross-linked to the nitro-cellulose by ultra violet light treatment (Spectrolinker XL-1500 UV crosslinker). The filters were then allowed to air dry at room temperature, and then pre-hybridised at 60° for one hour in a solution of 6xSSC with gentle agitation (for example using a Techne HB-1D hybridizer). Pre-hybridization blocks non-specific DNA binding sites on the filters.

In order to determine which colonies contained DNA inserts of interest the DNA crosslinked to the nitro-cellulose filter was hybridised with a radio-labelled ^{32}P-DNA probe prepared from HCPB PCR product of the pancreatic cDNA library (see above). About 50ng of DNA was labelled with 50µCi of ^{32}P-dCTP (~3000Ci/ mMol) using T7

DNA polymerase in a total volume of 50µl (Pharmacia T7 QUICKPRIME™ kit), and the reaction allowed to proceed for 15 minutes at 37°. The labelled probe was then heated to 95° for 2 minutes, to denature the double stranded DNA, immediately added to 10ml of 6xSSC at 60°, and this solution used to replace the pre-hybridisation solution on the filters. Incubation with gentle agitation was continued for about 3 hours at 60°. After this time the hybridisation solution was drained off, and the filters washed twice at 60° in 2xSSC for 15 minutes each time. Filters were then gently blotted dry, covered with cling film (SARAN^{TM} wrap or similar), and exposed against X-ray film (for example Kodak XOMAT-AR5^{TM}) over-night at room temperature. Following development of the film, colonies containing inserts of interest were identified as those which gave the strongest exposure (darkest spots) on the X-ray film. In this series of experiments about 15 % of the colonies gave positive hybridisation. From this 12 colonies were chosen for further screening. These colonies were picked from the duplicate filter, streaked and maintained on L-agar nutrient media containing 100µg/ml ampicillin, and grown in L-broth nutrient media containing 100µg/ml ampicillin.

The selected isolates were checked by PCR for inserts of the correct size, using primers BPT1 and BPB1, (SEQ ID NO: 1 and SEQ ID NO: 2), and for priming with an internal primer BPT2 (SEQ ID NO: 3) and BPB1. BPT2 is designed to prime at the end of the pro-sequence, prior to the start of the mature gene and to introduce an XbaI restriction site.

For PCR screening colonies of the selected isolates were picked and dispersed into 200µl of distilled water and heated at 100° for 10 minutes in a sealed Eppendorf tube. The suspensions were then centrifuged for 10 minutes in a microfuge to pellet cell debris, and 1µl of the supernatant used as the DNA template in PCR screening. Typically, 1µl of supernatant was mixed with 20pMols of two oligonucleotide primers, BPT1 and BPB1, or BPT2 and BPB1, dNTPs to a final concentration of 200µM, thermostable DNA polymerase reaction buffer, and 0.5U of thermostable DNA polymerase in a final volume of 20µl. The PCR incubation was carried out using 25 cycles of 94° for 1.5 minutes, 50° for 2 minutes, and 72° for 2 minutes, followed by a single incubation of 72° for 9.9 minutes at the end of the reaction.

The PCR products were analysed for DNA of the correct size (about 1250 base pairs from primers BPT1 to BPB1, and about 900 base pairs from primers BPT2 to BPB1, see **Figure 1**) by agarose gel electrophoresis. Ten of the twelve clones gave PCR DNA products of the correct size. Six of the ten clones were then taken for plasmid DNA preparation (using Qiagen Maxi kits, from 100ml of over-night culture at 37° in L-broth with 100µg/ml ampicillin). These plasmid DNA preparations were then sequenced over the region of PCR product insert using an USB Sequenase DNA sequencing kit, which incorporates bacteriophage T7 DNA polymerase. Each clone was sequenced using eight separate oligonucleotide primers, known as 676, 336, 337, 679, 677, 1280, 1279 and 1281 (SEQ ID NOs: 3 to 10). The positioning of the sequencing primers within the HCPB sequence is shown diagramatically in **Figure 2**, primers 336, 1279, 676, 1280, 677 and 1281 being 'forward', and 337 and 679 'backwards'.

Five of the six clones were found to have identical sequence (SEQ ID NO: 11) of 1263 base pairs between and including the SacI and XhoI restriction sites, and this sequence was used in further experiments. The translation of the DNA sequence into its polypeptide sequence is shown in SEQ ID NO: 12. The start of the mature protein sequence is amino acid residue 109. Amino acid numbered 14 marks the start of the putative pro-enzyme sequence. Only part of the enzyme secretion leader sequence (pre-sequence) is present in the cloned PCR generated DNA. The polypeptide sequence shows an aspartate residue at position 361, which when the whole sequence is aligned with other mammalian carboxypeptidase A and B sequences indicates a B type specificity (see amino acids numbered 255 by Catasus L, et al, Biochem J., 287, 299-303, 1992, and discussion). However, the cysteine residue at position 243 in the cloned sequence is not observed in other published human pancreatic carboxypeptidase B sequences, as highlighted by Yamamoto et al, in the Journal of Biological Chemistry, v267, 2575-2581, 1992, where she shows a gap in her sequence following the position numbered 244, when aligned with other mammalian pancreatic carboxypeptidase B amino acid sequences. Also shown on **Figure 2** are the approximate sites of the aspartate amino acid residue in the enzyme recognition site, and the cysteine residue at position 135 of the mature enzyme (position 243 in SEQ ID NO: 12).

One of the clones was deposited on 23-November-1994 with the National Collection of Industrial and Marine Bacteria Limited (23 St. Machar Drive, Aberdeen AB2 1RY, Scotland) and has the designation NCIMB 40694. The plasmid from this clone is known as pICI1698.

Reference Example 2

Expression of mature HCPB-(His)$_6$-c-Myc from E. coli

In order to achieve the expression of mature HCPB from E.coli the mature gene from pICI1698 was transferred into a plasmid vector which allows controlled secretion of protein products into the periplasm of the bacteria. This secretion vector, known as pICI266, in a bacterial host MSD522 suitable for controlled expression, has been deposited on 11 October 1993 with the National Collection of Industrial and Marine Bacteria Limited (Aberdeen AB2 1RY, Scotland) and has the designation NCIMB 40589. A plasmid map of pICI266 is shown in **Figure 3**. The plasmid has genes for tetracycline resistance and induction (TetA and TetR), an AraB operator and promoter sequence for inserted gene expression, and an AraC gene for expression control. The promoter sequence is followed by the PelB translation leader sequence which directs the polypeptide sequence following it to the periplasm. The site of gene cloning has several unique restriction sites and is followed by a phage T4 transcription terminator sequence. The DNA sequence in this region and the features for gene cloning are shown diagramatically in **Figure 4.**

For the cloning of the mature HCPB sequence into pICI266 it was decided to generate HCPB DNA by PCR, and to make some alterations to the codon usage at the start of the mature gene to introduce E.coli preferred codons. Also, to help with detection and purification of the expression construct a C-term peptide tag, known as (His)$_6$-c-myc was added to the enzyme. The tag consists of 6 histidines, a tri-peptide linker (EPE) and a peptide sequence (EQKLISEEDL) from c-myc which is recognised by the antibody 9E10 (as published by Evan et al., Mol Cell Biol, 5, 129-136, 1985, and available from Cambridge Research Biochemicals and other antibody suppliers). The C-terminus is completed by the addition of an Asparagine. The 6 histidine residues should allow the purification of the expressed protein on a metal chelate column (for example Ni-NTA

Agarose from Qiagen). In addition the PCR primers are used to introduce unique restriction sites at the 5' (FspI) and 3' (EcoRI) of the gene to facilitate the introduction of the PCR product into the expression vector. The sequence of the two primers, known as FSPTS1 and 6HIS9E10R1BS1, are shown in SEQ ID NOs: 13 and 14.

To generate a modified gene for cloning into pICI266, PCRs were set up using 100pMols of primers FSPTS1 and 6HIS9E10R1BS1 in the presence of approximately 5ng of pICI1698 DNA, dNTPs to a final concentration of 200μM, thermostable DNA polymerase reaction buffer, and 2.5U of thermostable DNA polymerase in a final volume of 100μl. The mixture was heated at 94° for 10 minutes prior to addition to the thermostable DNA enzyme, and the PCR incubation was carried out using 30 cycles of 94° for 1.5 minutes, 50° for 2 minutes, and 72° for 2 minutes, followed by a single incubation of 72° for 9.9 minutes at the end of the reaction. An aliquot of the PCR product was analysed for DNA of the correct size (about 1000 base pairs) by agarose gel electrophoresis and found to contain predominantly a band of the correct size. The remainder of the product from the reaction mix was purified and separated from excess reagents using a microconcentrator column (Centricon 100, Amicon), followed by DNA isolation by ethanol/sodium acetate precipitation, centrifugation, vacuum drying and re-suspension in distilled water. The isolated DNA was restriction digested with enzymes FspI and EcoRI, and a band of the correct size (about 1000 base pairs) purified.

pICI266 double stranded DNA, prepared using standard DNA technology (Qiagen plasmid kits or similar), was restriction digested with KpnI enzyme, being very careful to ensure complete digestion. The enzyme was then inactivated by heating at 65° for 10 minutes, and then cooling on ice. The 3' over-hang generated by the KpnI was then enzymatically digested by the addition of T4 DNA polymerase as recommended by the supplier (New England BioLabs), in the presence of dNTPs and incubation at 16° for 15 minutes. The reaction was stopped by inactivating the enzyme by heating at 70° for 15 minutes. The DNA product was purified from enzyme reaction contaminants using glass-milk, an aliquot checked for yield by agarose gel electrophoresis, and the remainder restriction digested with EcoRI enzyme. Again care was taken to ensure complete restriction digest. DNA of the correct size (about 5600 base pairs) was purified.

Aliquots of both restricted and purified DNA samples were checked for purity and concentration estimation using agarose gel electrophoresis compared with known standards. From these estimates ligation mixes were prepared to clone the HCPB gene into the vector, using a molar ratio of about 1 vector to 2.5 insert (1 pICI266 to 2.5 HCPB PCR product), and a final DNA concentration of about 2.5ng/µl, in the presence of T4 DNA ligase, 1 mM ATP and enzyme buffer, using conditions suitable for the ligation of blunt ended DNA (FspI to T4 DNA polymerase treated KpnI).

Following the ligation reaction the DNA mixture was used to transform E.coli strain DH5α. Cell aliquots were plated on L-agar nutrient media containing 10µg/ml tetracycline as selection for plasmid vector, and incubated over-night at 37°. Colonies containing plasmids with inserts of interest were identified by hybridisation as described in Reference Example 1 except that the 12 colonies chosen for further screening were treated as follows.

These colonies were picked from the duplicate filter, streaked and maintained on L-agar nutrient media containing 10µg/ml tetracycline, and grown in L-broth nutrient media containing 10µg/ml tetracycline.

The selected isolates were checked by PCR for inserts of the correct size, using primers FSPTS1 and 6HIS9E10R1BS1, (SEQ ID NO: 13 and SEQ ID NO: 14), and for priming with an internal primer BPB2 (SEQ ID NO: 6) and FSPT1. BPB2 is designed to prime within the mature gene and generate a fragment of about 430 base pairs.

For PCR screening colonies of the selected isolates were picked and dispersed into 200µl of distilled water and heated at 100° for 10 minutes in a sealed tube. The suspensions were then centrifuged for 10 minutes in a microfuge to pellet cell debris, and 1 µl of the supernatant used as the DNA template in PCR screening. Typically, 1 µl of supernatant was mixed with 20 pMols of two oligonucleotide primers, FSPT1 and 6HIS9E10R1BS1, or FSPT1 and BPB2, dNTPs to a final concentration of 200 µM, PCR Buffer, and 0.5 U of thermostable DNA polymerase in a final volume of 20 µl. The PCR incubation was carried out using 25 cycles of 94° for 1.5 minutes, 50° for 2 minutes, and 72° for 2 minutes, followed by a single incubation of 72° for 9.9 minutes at the end of the reaction.

The PCR products were analysed for DNA of the correct size (about 1000 base pairs from primers FSPTS1 to 6HIS9E10R1BS1, and about 430 base pairs from primers FSPTS1 to BPB2) by agarose gel electrophoresis. All twelve clones gave PCR DNA products of the correct size. Six of the clones were then taken for plasmid DNA preparation (using Qiagen MAXI™ kits, from 100 ml of over-night culture at 37° in L-broth with 10 μg/ml tetracycline). These plasmid DNA preparations were then sequenced over the region of PCR product insert using a DNA sequencing kit (USB SEQUENASE™), which incorporates bacteriophage T7 DNA polymerase. Alternatively the DNA was sequenced using an automated DNA sequencer. The clones were sequenced using several separate oligonucleotide primers. Three of the primers, known as 1504, 1590 and 1731, were used to check the cloning junctions between the expression vector and the inserted gene (SEQ ID NOs: 15, 16 and 17), as well as giving sequence data from the start and end of the inserted gene. Other primers, including those known as 679, 677, 1802, and 1280 (SEQ ID NOs: 6, 7, 18 and 8) were used to confirm the remainder of the inserted gene sequence. This plasmid containing the modified mature HCPB gene is known as pICI1712.

To obtain controlled expression of the modified HCPB the pICI1712 plasmid DNA was transformed into calcium chloride transformation competent E.coli expression strains. Included amongst these strains were a number which were incapable of growing with arabinose as the major carbon source, and were chromosome deleted for the arabinose (Ara) operon. A preferred strain is known as MSD213 (strain MC1000 of Casadaban et al, Journal of Molecular Biology, 138, 179-208, 1980), and has the partial genotype, F- Ara Δ (Ara-Leu) ΔLacX74 GalV GalK StrR. Another preferred strain is known as MSD525 (strain MC1061) and has the genotype, AraD139 Δ(Ara Leu)7697 ΔLac74 GalU HsdR RpsL. E.coli strains of similar genotype, suitable for controlled expression of genes from the AraB promoter in plasmid pICI266, may be obtained from The E.coli Genetic Stock Centre, Department of Biology, Yale University, CT, USA. Selection for transformation was on L-agar nutrient media containing 10μg/ml tetracycline, over night at 37°. Single colonies were picked from the transformation plates, purified by streaking and maintained on L-agar nutrient media containing 10μg/ml tetracycline, and grown in L-broth nutrient media containing 10μg/ml tetracycline.

All pICI1712 transformed expression strains were treated in the same manner to test for expression of the cloned HCPB gene.

1. A single colony was used to inoculate 10 ml of L-broth nutrient media containing 10 µg/ml tetracycline in a 25 ml universal container, and incubated over night at 37° with shaking.

2. 75 ml of L-broth nutrient media containing 10 µg/ml tetracycline pre-warmed to 37° in a 250 ml conical flask was inoculated with 0.75 ml (1 % v/v) of the over-night culture. Incubation was continued at 37° with shaking, and growth monitored by light absorbance at 540 nm. Induction of cloned protein expression was required during exponential growth of the culture, and this was taken as between 0.4 and 0.6 O.D. at 540 nm, and generally took 90 to 150 minutes from inoculation.

3. When the cells had reached the required optical density the cultures were allowed to cool to approximately 30° by placing the flasks at room temperature for 30 minutes. Arabinose was then added to a final concentration of 1 % (w/v), and incubation continued at 30° with shaking for 4 to 6 hours.

4. After incubation a final optical density measurement is taken, and the cells were harvested by centrifugation. The final O.D. measurement is used to calculate the the volume of protein acrylamide gel (Laemmli) loading buffer that is used to resuspend the cell pellet. For O.D. less than 1 a volume of 10µl is used for each 0.1 O.D. unit, and for an O.D. greater than 1 a volume of 15 µl is used for each 0.1 O.D. unit.

5. Following re-suspension the samples were denatured by heating at 100° for 10 minutes, and then centrifuged to separate the viscous cell debris from the supernatant. Expression samples, usually 20 µl of the supernatant, typically were loaded onto 17 % SDS acrylamide gels for electrophoretic separation of the proteins. Duplicate gels were generally prepared so that one could be stained for total protein (using Coomassie or similar stain and standard conditions), and the other could be processed to indicate specific products using Western analysis.

For Western analysis proteins in the run gel were transferred to nylon membrane (PROBLOT™, Applied Biosystems for example), using a semi-dry electrophoresis blotting apparatus (BioRad or similar). Before and during processing care was taken to ensure that the membrane remained damp. After transfer of the proteins from the gel,

further binding was blocked with a solution of 5 % low fat milk powder (MARVEL™ or similar) in PBS at room temperature with gentle agitation for 5 hours. The membrane was then washed 3 times at room temperature with gentle agitation for 5 minutes each time in PBS containing 0.05 % Tween™ 20. The washed membrane was then incubated with the primary antibody, monoclonal 9E10 mouse anti-c-myc peptide (see above), at a suitable dilution (typically 1 in 10,000 for ascites or 1 in 40 for hybridoma culture supernatant) in PBS containing 0.05 % Tween™ 20 and 0.5 % low fat milk powder, at room temperature with gentle agitation over night. The membrane was then washed 3 times at room temperature with gentle agitation for at least 5 minutes each time in PBS containing 0.05 % TWEEN™ 20. The washed membrane was then incubated with the secondary antibody, horseradish peroxidase labelled anti-mouse IgG (typically raised in goat, such as A4416 from Sigma), at a suitable dilution (typically 1 in 10,000) in PBS containing 0.05 % TWEEN™ 20 and 0.5 % low fat milk powder, at room temperature with gentle agitation for at least three hours. The membrane was then washed 3 times at room temperature with gentle agitation for at least 10 minutes each time in PBS containing 0.05 % TWEEN™ 20. The membrane was then processed using a chemiluminescence Western detection kit (Amersham ECL™) and exposed against film (Amersham HYPERFILM ECL™) for 30 seconds in the first instance, and then for appropriate times to give a clear image of the expressed protein bands. Other methods of similar sensitivity for the detection of peroxidase labelled proteins on membranes may be used.

Good expression of the cloned tagged HCPB in pICI266 (pICI1712) was demonstrated in E.coli strains MSD213 and MSD525 by the Coomassie stained gels showing an additional strong protein band at about 35,000 Daltons when compared to vector (pICI266) alone clones, and a band of the same size giving a strong signal by Western analysis detection of the c-myc peptide tag.

Reference Example 3

Purification of native HCPB

A system has been determined for the initial purification of the native and the different mutant enzymes via two routes.

The preferred route is described first. Recombinant E.coli cell paste containing the recombinant enzyme was taken from storage at -70° and allowed to thaw. The weight of cell paste was measured in grams and the paste resuspended with the addition of Buffer A to a volume equal to the initial weight of the cell paste. The cell suspension was incubated at room temperature for 20 minutes with occasional gentle mixing before an equal volume of distilled water was added and thoroughly mixed in. The cell suspension was again incubated at room temperature for 20 minutes with occasional gentle mixing. The resulting crude osmotic shockate was clarified by centrifugation at 98000 x g for 90 minutes at 4° after which the supernatant was decanted off from the pelleted insoluble fraction. Deoxyribonuclease 1 was added to the supernatant to a final concentration of 0.1 mg/ml The mixture was incubated at room temperature, with continuous shaking, until the vicosity was reduced enough for it to be loaded on to a Carboxypeptidase Inhibitor CNBr activated affinity column (CNBr activated SEPHAROSE™ 4B from Pharmacia) and carboxypeptidase inhibitor from potato tuber (c-0279,Sigma). The supernatant was adjusted to pH8.0 and loaded on to the affinity column, pre-equilibrated with 10 mM TRIS-HCl, 500 mM sodium chloride, pH 8.0. After loading the supernatant the column was washed until the absorbance of the flow through was back to baseline before the bound material was eluted from the column by Elution Buffer. The eluted fractions were frozen at -20° whilst those containing the recombinant carboxypeptidase were determined by Western blot analysis using an anti- c-myc tag antibody (9E10), followed by an anti-mouse -horse radish peroxidase conjugate (A-9044, Sigma) that gave a colour reaction with exposure to 4-chloronaphthol and hydrogen peroxide.

Fractions containing the recombinant carboxypeptidase B were pooled, concentrated and the pH adjusted to pH 7.5 before being snap-frozen and stored at -20°. Further purification of the pooled sample, utilising known methods such as ion exchange and gel permeation chromatography may performed if required.

The second route involves the total lysis of the E.coli cells as opposed to a periplasmic shock, as used in the preferred route.

Recombinant E.coli cell paste containing the recombinant enzyme was taken and resuspended in Lysis Buffer. Lysozyme was added to a concentration of 1mg/ml and at the same time lithium dodecyl sulphate (LDS) was added (80µl of a 25 % solution per

25ml of suspension). The suspension was incubated on ice for 30minutes with occasional shaking, followed by the addition deoxyribonuclease 1 to a concentration of 1mg/ml and again the suspension was incubated on ice for 30 minutes with occasion shaking. The suspension was subsequently divided in to 200ml volumes and sonicated to complete the disruption of the cells for 10 x 30 sec bursts with 30sec intervals between bursts. Sonicated suspensions were centrifuged at 98,000x g for 90 minutes at 4° after which the supernatant was decanted off from the pelleted insoluble fraction. The supernatant was adjusted to pH 8.0 and loaded on to the affinity column, pre-equilibrated with 10 mM TRIS-HCl, 500 mM sodium chloride, pH 8.0. After loading the supernatant the column was washed until the absorbance of the flow through was back to baseline before the bound material was eluted from the column by Elution Buffer. The eluted fractions were frozen at -20° whilst those containing the recombinant carboxypeptidase were determined by Western blot analysis using an anti- c-myc tag antibody (9E10), followed by an anti-mouse -horse radish peroxidase conjugate (A-9044, Sigma) that gave a colour reaction with exposure to 4-chloronaphthol and hydrogen peroxide. Fractions containing the recombinant carboxypeptidase B were pooled, concentrated and the pH adjusted to pH 7.5 before being snap-frozen and stored at -20°. Further purification of the pooled sample. utilising known methods such as ion exchange and gel permeation chromatography may performed if required.

Samples of the pooled material from both routes, analysed by SDS-PAGE and Coomassie stained nitrocellulose blot provided Coomassie stained bands at the correct molecular weight for the recombinant carboxypeptidase B's. These bands sequenced by an automated protein/peptide sequencer using the Edman degradation technique gave positive matches for the particular recombinant carboxypeptidase B being purified.

Reference Example 4

Expression of mature HCPB from COS cells by co-secretion of the pro sequence

A gene encoding preHCPB was generated by PCR from pICI1698 (Reference Example 1). The PCR was set up with template pICI1689 (10ng) and oligos SEQ ID NO: 19 and SEQ ID NO: 20 (100pMoles of each) in buffer (100µl) containing 10 mM Tris-HCl (pH8.3). 50 mM KCL, 1.5 mM $MgCl_2$, 0.125 mM each of dATP, dCTP, dGTP and dTTP

and 2.5u thermostable DNA polymerase. The reaction was overlaid with mineral oil (100µl) and incubated at 94° for 1 min, 53° for 1 min and 72° for 2.5 min for 25 cycles, plus 10 min at 72°. The PCR product of 985bp was isolated by electrophoresis on a 1 % agarose (Agarose type I, Sigma A-6013) gel followed by excision of the band from the gel and isolation of the DNA fragment.

The preHCPB gene was digested for 1h at 37° with EcoRI and HindIII in a 100µl reaction containing 100 mM Tris-HCl (pH 7.5), 10 mM magnesium chloride, 50 mM NaCl, 0.025 % TRITON™ X-100, and 25u each of HindIII and EcoRI (New England Biolabs). The digested fragment was purified and cloned into pBluescript (Stratagene Cloning Systems).

pBluescript KS+ DNA (5µg) was digested to completion with EcoRI and HindIII (25u each) in a 100µl reaction as described above. Calf-intestinal alkaline phosphatase (1µl; New England Biolabs, 10u/µl) was the added to the digested plasmid to remove 5' phosphate groups and incubation continued at 37° for a further 30 minutes. Phosphatase activity was destroyed by incubation at 70° for 10 minutes. The EcoRI-HindIII cut plasmid was purified from an agarose gel as described above. The EcoRI-HindIII digested preHCPB gene (50 ng) was ligated with the above cut plasmid DNA in 20 µl of a solution containing 30 mM Tris-Hcl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 50 µg/ml BSA and 400u T4 DNA ligase at 25° for 4h. A 1µl aliquot of the reaction was used to transform 20µl of competent E. coli DH5α cells. Transformed cells were plated onto L-agar plus 100µg/ml Ampicillin. Potential preHCPB clones were identified by PCR. Each clone was subjected to PCR as described above for preparation of the preHCPB gene except that the mix with the cells was incubated at 94° (hot start procedure) for 5 min prior to 25 cycles of PCR and oligos SEQ ID NOS 21 and 22 replace oligos SEQ ID NOS: 19 and 20. A sample (10µl) of the PCR reaction was analysed by electrophoresis on a 1 % agarose gel. Clones containing the preHCPB gene were identified by the presence of a 1.2kb PCR product. Clones producing the 1.2kb PCR product were used for large scale plasmid DNA preparation and the sequence of the insert confirmed by DNA sequence analysis. The plasmid containing the preHCPB gene in pBluescript was named pMF15.

To generate vectors capable of expressing HCPB in eukaryotic cells the GS-System(TM) system (Celltech Biologics) was used (WO 87/04462, WO 89/01036,

WO 86/05807 and WO 89/10404). The procedure requires cloning the preHCPB gene into the HindIII-EcoRI region of vector pEE12 [this vector is similar to pSV2.GS described in Bebbington et al. (1992) Bio/Technology 10, 169-175, with a number of restriction sites originally present in pSV2.GS removed by site-directed mutagenesis to provide unique sites in the multi-linker region]. To construct the expression vector, plasmids pEE12 and pMF15 were digested with EcoRI and HindIII as described above. The appropriate vector (from pEE12) and insert (from pMF15) from each digest were isloated from a 1 % agarose gel and ligated together and used to transform competent DH5 α cells. The transformed cells were were plated onto L agar plus ampicillin (100μg/ml). Colonies were screened by PCR as described above, with oligos which prime within the CMV promoter (SEQ ID NO: 23) and in the HCPB gene (SEQ ID NO: 24). Clones producing a 1.365kb PCR product were used for large scale plasmid DNA preparation and the sequence of the insert confirmed by DNA sequence analysis. The plasmid containing the preHCPB sequence in pEE12 was named pMF48.

A second eukaryotic expression plasmid, pEE12 containing the prepro sequence of preproHCPB was prepared as described above. First, a gene for preproHCPB was prepared by PCR using as template pICI1689 and oligos SEQ ID NOS: 19 and 25 to give a 1270bp PCR product. The gene was digested with EcoRI and HindIII and cloned initially into pBluescript KS+ to give pMF18. Next, oligos SEQ ID NOS: 25 and 26 were used in a PCR to isolate a gene for the prepro sequence from pMF18. In this case the PCR was performed with a hot start procedure by first incubating the mix without thermostable DNA polymerase for 5 min at 94°. Thermostable DNA polymerase (2.5u) was then added and the PCR continued through the 25 cycles as described above. The 356bp fragment was purified then digested with EcoRI and HindIII and cloned into pBluescript to give pMF66 and subsequently into pEE12 (screening by PCR with SEQ ID NOS 25 and 26) to give pMF67.

For expression in eukaryotic cells, vectors containing genes capable of expressing HCPB (amino acid residues 109 to 415 of SEQ ID NO: 12) and the pro sequence (amino acid residues 14 to 108 of SEQ ID NO: 12) were cotransfected into COS-7 cells. COS cells are an African green monkey kidney cell line, CV-1, transformed with an origin-defective SV40 virus and have been widely used for short-term transient expression

of a variety of proteins because of their capacity to replicate circular plasmids containing an SV40 origin of replication to very high copy number. There are two widely available COS cell clones, COS-1 and COS-7. The basic methodology for transfection of COS cells is described by Bebbington in Methods: A Companion to Methods in Enzymology (1991) 2, p. 141. For expression of HCPB, the plasmid vectors pMF48 and pMF67 (2μg of each) were used to transfect the COS-7 cells (2 X 10^5) in a six-well culture plate in 2ml DMEM containing 10 % heat inactivated FCS by a method known as lipofection - cationic lipid-mediated delivery of polynucleotides [Felgner et al. in Methods: A Companion to Methods in Enzymology (1993) 5, 67-75]. The cells were incubated at 37° in a CO_2 incubator for 20h. The mix of plasmid DNA in serum-free medium (200μl) was mixed gently with LIPOFECTIN™ reagent (12μl) and incubated at ambient temperature for 15min. The cells were washed with serum-free medium (2ml). Serum-free medium (600μl) was added to the DNA/LIPOFECTIN™ and the mix overlaid onto the cells which were incubated at 37° for 6h in a CO_2 incubator. The DNA containing medium was replaced with normal DMEM containing 10% FCS and the cells incubated as before for 72h. Cell supernatants (250μl) were analysed for HCPB activity against Hipp-Arg (5h assay, in a total volume of 500μl) essentially as described in Example 5. COS cell supernatants which had been treated with LIPOFECTIN™ reagent, but without plasmid DNA, hydrolysed 1.2 % of the substrate, whereas the COS cell supernatants transfected with the mix of plasmids expressing preHCPB and prepro sequence hydrolysed 61 % of the Hipp-Arg substrate. COS cells transfected with only the preHCPB plasmid hydrolysed Hipp-Arg at the level seen for COS cells which had been treated with LIPOFECTIN™ reagent alone.

Reference Example 5

Preparation of pMF133

This example describes the preparation of a plasmid for expression of [G251T,D253K]HCPB-His_6 -cMyc.

Plasmids pMF48 (10 μg; described in Reference Example 4) and pMC46.4.1 (10 μg; described in Example 14 of International Patent Application WO 97/07769 Zeneca Ltd, published 6 March, 1997) were digested separately to completion with XmaI (10

units; New England Biolabs) in a 100 µl reaction containing 10 mM Bis Tris-Propane-HCl (pH 7.0), 10 mM magnesium chloride, 1 mM DTT at 37°. The cut plasmids were purified by electrophoresis on a 1 % agarose gel (Agarose type I, Sigma A-6013 followed by excision of the band from the gel and isolation of the DNA fragment. The Xma cut fragments were further digested with EcoRI (140 units; New England Biolabs) in a 100 µl reaction containing 100 mM Tris-HCl (pH 7.5), 10 mM magnesium chloride, 50 mM NaCl, 0.025 % Triton™ X-100 at 37° for 1 h. The digested vector fragment from pMF48 (7833 bp) and insert fragment (271 bp) from pMC46.4.1 were isolated from a 1 % agarose gel as described above. The pMF48 vector fragment and pMF46.4.1 insert fragment were ligated in 20 µl of a solution containing 30 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 50 µg/ml BSA and 400u T4 DNA ligase at 25° for 4 h. A 1 µl aliquot of the reaction was used to transform 20 µl of competent E. coli DH5α cells. Transformed cells were plated onto L-agar plus 100 µg/ml Ampicillin. Clones containing a pre[G251T,D253K]HCPB-His_6 -cMyc gene were identified by PCR as described for the preHCPB gene in Reference Example 4, by use of oligos SEQ ID Nos: 23 and 24. Clones producing the 1.365 kb PCR product were used for large scale plasmid DNA preparation and the sequence of the insert confirmed by DNA sequence analysis. The plasmid containing the pre[G251T,D253K]HCPB-His_6 -cMyc gene in pEE12 was named pMF133.

Reference Example 6

Preparation of IgG3-pBSIIKS+

This example describes the preparation of a vector containing a gene for the human IgG3 heavy chain constant and hinge region.

A gene containing the sequence shown in SEQ ID NO: 39 was prepared by PCR by a method similar to that described by Jayaraman et al. (1991) Proc. Natl. Acad. Sci USA 88, 4084-4088.

The gene was made in two parts, known as IgG3A and IgG3B. These were cloned separately into the SacI and XmaI sites of pBluescript KS+ (Stratagene Cloning Systems) to give vectors IgG3A-pBSIIKS+ clone A7 and IgG3B-pBSIIKS+ clone B17 respectively. IgG3A was made to extend past the PmaCI restriction site (CACGTG at positions 334-339 in SEQ ID NO: 39). Similarly, IgG3B was made such that the 5` end of the sequence was

upstream of the PmaCI restriction site. To obtain the desired IgG3 gene sequence, the intermediate IgG3A and IgG3B vectors were cut with AflIII and PmaCI. The vector fragment (2823bp) from IgG3A-pBSIIKS+ clone A7, and insert fragment from IgG3B-pBSIIKS+ clone B17 (666bp) were isolated by electrophoresis in a 1 % agarose gel and purified. The fragments were ligated and the ligation mix used to transform E. coli strain DH5α. Clones containing the required gene were identified by digestion of isolated DNA with SacI and XmaI to give a 520bp fragment. The sequence of the insert was confirmed by DNA sequence analysis and clone number F3 was designated IgG3-pBSIIKS+.

Reference Example 7

Preparation of plasmid pNG3-VKss-806.077HuVK4-HuCK-Neo

A synthetic DNA sequence of SEQ ID NO: 49 was prepared using PCR by a method similar to that described by Jayaraman et al. (1991) Proc. Natl. Acad. Sci USA 88, 4084-4088 and cloned into pNG-VKss-HuCK-Neo (NCIMB deposit no. 40799, deposited 11-Apr-96 at National Collection of Industrial and Marine Bacteria Limited, 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland). To achieve this, the synthetic sequence of SEQ ID NO: 49 was digested with SacII and XhoI restriction enzymes and cloned into similarly digested pNG-VKss-HuCK-Neo. The plasmid vector so produced was named pNG3-VKss-806.077HuVK1-HuCK-Neo. Plasmid pNG3-VKss-806.077HuVK1-HuCK-Neo was used as the template for two PCR reactions. General reaction conditions were as follows.

To 5 µl of the cDNA reaction was added 5 µl dNTPs (2.5 mM), 5 µl 10x Enzyme Buffer, 1 µl of 25 pM/µl back primer, 1 µl of 25 pM/µl forward primer, 0.5 µl thermostable DNA polymerase and DEPC-treated water to obtain a volume of 50 µl. The PCR conditions were set for 15 cycles at 94° for 90 s; 55° for 60 s; 72° for 120 s, ending the last cycle with a further 72° for 10 min incubation.

Reaction A used the synthetic oligonucleotide sequence primers SEQ ID NOS: 50 and 51 and reaction B the synthetic oligonucleotide sequence primers SEQ ID NOS: 52 and 53. The products of these PCR reactions (A and B) were fragments of length 535 base pairs and 205 base pairs respectively. These reaction products were run on a 2 % agarose gel and separated from any background products. Bands of the expected size were excised from the gel and recovered. Mixtures of varying amounts of the products A and B were made and PCR

reactions performed using the synthetic oligonucleotides SEQ ID NOS: 50 and 52. The resulting product (ca.700 base pairs) was digested with the restriction enzymes SacII and XhoI and the cleavage products separated on a 2 % agarose gel. The band of the expected 310 base pairs size was excised from the gel and recovered. This fragment was then ligated into the vector pNG3-806.077HuVK1-HuVK-Neo vector (which had been previously cut with the restriction enzymes SacII/XhoI and subsequently isolated) and thus created HuVK4 DNA sequence (SEQ ID NO: 45) in the vector pNG3-Vkss-806.077HuVK4-HuCK-Neo.

Example 1

Expression of mature HCPB from COS cells by co-secretion of pro-L sequence

The procedure described in Reference Example 4 was repeated but pMF67, containing the gene for the prepro sequence, was replaced by plasmid pMF161 containing a gene for a pro-L modified prepro sequence. The pro-L sequence is the natural pro sequence with a C-terminal leucine residue. The amino acid sequence of pro-L is shown in SEQ ID NO: 27.

Plasmid pMF161 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 28. The 359bp fragment was cloned into pBluescript to give pMF141 and subsequently into pEE12 (Screening by PCR with SEQ ID NOS: 7 and 28) to give pMF161.

Cell supernatants diluted 1 in 80 (250µl) were analysed for HCPB activity against Hipp-Arg (5h assay) as described in Example 5. The result is shown in Example 5 and is expressed as the percentage conversion of substrate into product at 37°.

Example 2

Expression of mature HCPB from COS cells by co-secretion of pro-KDEL sequence

The procedure described in Reference Example 4 was repeated but with pMF67, containing the gene for the prepro sequence, was replaced by plasmid pMF164 containing a gene for a pro-KDEL modified prepro sequence. The pro-KDEL sequence is the natural pro sequence with a C-terminal KDEL tetrapeptide residue. The amino acid sequence of pro-KDEL is shown in SEQ ID NO: 29.

Plasmid pMF164 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 30. The 365bp fragment was cloned into pBluescript to give pMF149 and subsequently into pEE12 (screening by PCR with SEQ ID NOS: 7 and 30) to give pMF164.

Cell supernatants diluted 1 in 80 (250µl) were analysed for HCPB activity against Hipp-Arg (5h assay) as described in Example 5. The result is shown in Example 5 and is expressed as the percentage conversion of substrate into product at 37°.

Example 3

Expression of mature HCPB from COS cells by co-secretion of pro-KKAA sequence

The procedure described in Reference Example 4 was repeated but with pMF67, containing the gene for the prepro sequence, was replaced by plasmid pMF165 containing a gene for a pro-KKAA modified prepro sequence. The pro-KKAA sequence is the natural pro sequence with a C-terminal KKAA tetrapeptide residue. The amino acid sequence of pro-KKAA is shown in SEQ ID NO: 31.

Plasmid pMF165 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 32. The 365bp fragment was cloned into pBluescript to give pMF145 and subsequently into pEE12 (Screening by PCR with SEQ ID NOS: 7 and 32) to give pMF165.

Cell supernatants diluted 1 in 80 (250µl) were analysed for HCPB activity against Hipp-Arg (5h assay) as described in Example 5. The result is shown in Example 5 and is expressed as the percentage conversion of substrate into product at 37°.

Example 4

Expression of mature HCPB from COS cells by co-secretion of pro-SDYQRL sequence

The procedure described in Reference Example 4 was repeated but with pMF67, containing the gene for the prepro sequence, was replaced by plasmid pMF166 containing a gene for a pro-SDYQRL modified prepro sequence. The pro-SDYQRL sequence is the natural pro sequence with a C-terminal SDYQRL hexapeptide residue. The amino acid sequence of pro-SDYQRL is shown in SEQ ID NO: 33.

Plasmid pMF166 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 34. The 371bp fragment was cloned into pBluescript to give pMF148 and subsequently into pEE12 (Screening by PCR with SEQ ID NOS: 7 and 34) to give pMF166.

Cell supernatants diluted 1 in 80 (250µl) were analysed for HCPB activity against Hipp-Arg (5h assay) as described in Example 5. The result is shown in Example 5 and is expressed as the percentage conversion of substrate into product at 37°.

Example 5

Enzymic activity of recombinant HCPB against Hipp-Arg.

Purified human CPB, produced as described in Reference Example 3, or COS cell supernatant (from Reference Example 4 and Examples 1 to 4) was assayed for its ability to convert hippuryl-L-arginine (Hipp-Arg; Sigma) to hippuric acid using a HPLC assay.

The reaction mixture contained either purified human CPB, or COS cell supernatant diluted 1 in 8 or 1 in 80, and 0.5 mM Hipp-Arg in 0.025 M Tris-HCL, pH 7.5 (250 µl total volume). Samples were incubated for 5 hr at 37°. The reactions were terminated by the addition of 250 µl of 40 % methanol, 60 % Phosphate Buffer, 0.2 % trifluoroacetic acid and the amount of hippuric acid generated was quantified by HPLC.

HPLC analysis was carried out using a Hewlett Packard 1090 Series 11 (with diode array) HPLC system. Samples (50 µl) were injected onto a HICHROM Hi-RPB™ column (25 cm) and separated using a mobile phase of 20 % methanol, 80 % Phosphate Buffer at a flow rate of 1ml/min. The amount of product (hippuric acid, detected at 230nm) produced was determined from calibration curves generated with known amounts of hippuric acid (Sigma-H6375). The results are expressed as the percentage conversion of substrate into product in 5 hr at 37°.

The data demonstrate that co-expression of HCPB in the presence of pro sequences with additional C-terminal amino-acid residues is about 2 to 15 fold higher than when expressed with an unmodified pro sequence.

Table - HCPB expression from COS cells

		% hydrolysis of Hipp-Arg	
Protein expressed	Plasmid	1:8 dilution	1:80 dilution
mature HCPB	pMF48	1.0	1.4
+ pro	pMF48+pMF67	39.8	4.3
+ pro L	pMF48+pMF161	100	39.6
+ pro KDEL	pMF48+pMF164	85.3	9.9
+ pro KKAA	pMF48+pMF165	100	59.9
+ pro SDYQRL	pMF48+pMF166	100	40.5

HCPB (0.625 ng) gave 10 % hydrolysis in a similar 5 h assay

Example 6

Expression of murine A5B7 F(ab')$_2$-(HCPB)$_2$ fusion protein from COS cells by co-secretion of pro-L sequence

The procedure described in Reference Example 13 of International Patent Application WO 97/07769 (Zeneca Ltd, published 6 March, 1997) was repeated but with pMF67, containing the gene for the prepro sequence of HCPB, replaced by plasmid pMF161 containing a gene for a pro-L modified prepro sequence. The pro-L sequence is the natural pro sequence with a C-terminal leucine residue. The amino acid sequence of pro-L is shown in SEQ ID NO: 27.

Plasmid pMF161 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 28. The 359 bp fragment was cloned into pBluescript to give pMF141 and subsequently into pEE12 (Screening by PCR with SEQ ID NOS: 7 and 28) to give pMF161.

Cell supernatants diluted 1 in 80 (250 µl) were analysed for HCPB activity against Hipp-Arg (5 h assay) as described in Example 9. The result is shown in Example 9 and is expressed as the percentage conversion of substrate into product at 37°.

Example 7

Expression of murine A5B7 F(ab')$_2$-(HCPB)$_2$ fusion protein from COS cells by co-secretion of pro-KKAA sequence

The procedure described in Reference Example 13 of International Patent Application WO 97/07769 (Zeneca Ltd, published 6 March, 1997) was repeated but with pMF67, containing the gene for the prepro sequence of HCPB, replaced by plasmid pMF165 containing a gene for a pro-KKAA modified prepro sequence. The pro-KKAA sequence is the natural pro sequence with a C-terminal KKAA tetrapeptide residue. The amino acid sequence of pro-KKAA is shown in SEQ ID NO: 31.

Plasmid pMF165 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 32. The 365bp fragment was cloned into pBluescript to give pMF145 and subsequently into pEE12 (Screening by PCR with SEQ ID NOS: 7 and 32) to give pMF165.

Cell supernatants diluted 1 in 80 (250 µl) were analysed for HCPB activity against Hipp-Arg (5 h assay) as described in Example 9. The result is shown in Example 9 and is expressed as the percentage conversion of substrate into product at 37°.

Example 8

Expression of murine A5B7 F(ab')$_2$-(HCPB)$_2$ fusion protein from COS cells by co-secretion of pro-SDYQRL sequence

The procedure described in Reference Example 13 of International Patent Application WO 97/07769 (Zeneca Ltd, published 6 March, 1997) was repeated but with pMF67, containing the gene for the prepro sequence of HCPB, replaced by plasmid pMF166 containing a gene for a pro-SDYQRL modified prepro sequence. The pro-SDYQRL sequence is the natural pro sequence with a C-terminal SDYQRL hexapeptide residue. The amino acid sequence of pro-SDYQRL is shown in SEQ ID NO: 33.

Plasmid pMF166 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 34. The 371 bp fragment was cloned into pBluescript to give pMF148 and subsequently into pEE12 (screening by PCR with SEQ ID NOS: 7 and 34) to give pMF166.

Cell supernatants diluted 1 in 80 (250 µl) were analysed for HCPB activity against Hipp-Arg (5 h assay) as described in Example 9. The result is shown in Example 9 and is expressed as the percentage conversion of substrate into product at 37°.

Example 9

Enzymic activity and antigen binding activity of recombinant murine A5B7 F(ab')$_2$-(HCPB)$_2$ fusion protein

COS cell supernatant (from Reference Example 13 of International Patent Application WO 97/07769, Zeneca Ltd, published 6 March, 1997) and Examples 6 to 8) was assayed for its ability to convert hippuryl-L-arginine (Hipp-Arg; Sigma) to hippuric acid using a HPLC assay.

The reaction mixture contained either purified human CPB, or COS cell supernatant diluted 1 in 8 or 1 in 80, and 0.5 mM Hipp-Arg in 0.025 M Tris-HCL, pH 7.5 (250 µl total volume). Samples were incubated for 5 h at 37°. The reactions were terminated by the addition of 250 µl of 40 % methanol, 60 % Phosphate Buffer, 0.2 % trifluoroacetic acid and the amount of hippuric acid generated was quantified by HPLC.

HPLC analysis was carried out using a suitable HPLC system (Hewlett Packard 1090 Series 11 with diode array). Samples (50 µl) were injected onto a HICHROM Hi-RPB™ column (25 cm) and separated using a mobile phase of 20 % methanol, 80 % Phosphate Buffer at a flow rate of 1 ml/min. The amount of product (hippuric acid. detected at 230 nm) produced was determined from calibration curves generated with known amounts of hippuric acid (Sigma-H6375). The results are expressed as the percentage conversion of substrate into product in 5 h at 37°.

The data demonstrate that co-expression of A5B7 F(ab')$_2$-(HCPB)$_2$ fusion protein in the presence of pro sequences with additional C-terminal amino-acid residues is enhanced than when expressed with an unmodified pro sequence.

Table - Enzyme activity of A5B7 F(ab')$_2$-(HCPB)$_2$ expression from COS cells

		% hydrolysis of Hipp-Arg	
Protein expressed	Plasmid	1:8 dilution	1:80 dilution
none		2.1	0
F(ab')$_2$-(HCPB)$_2$	pMF53	2.2	0
+ pro	pMF53+pMF67	100	33.7
+ pro L	pMF53+pMF161	98.5	70.4
+ pro KKAA	pMF53+pMF165	98.2	51.3
+ pro SDYQRL	pMF53+pMF166	99.3	64.4

COS cell supernatant were also assayed for antigen binding activity in a CEA ELISA assay performed essentially as described in Reference Example 13 of International Patent Application WO 97/07769, Zeneca Ltd, published 6 March, 1997. The results are expressed as concentrations of fusion protein present in the supernatant. The amount of A5B7 F(ab')$_2$-(HCPB)$_2$ fusion protein was determined from calibration curves generated with known amounts of A5B7 F(ab')$_2$. Preparation of A5B7 F(ab')$_2$ is described in Reference Example 5 of International Patent Application WO 96/20011, Zeneca Ltd, published 4 July, 1996.

The data demonstrate that co-expression of A5B7 F(ab')$_2$-(HCPB)$_2$ fusion protein in the presence of pro sequences with additional C-terminal amino-acid residues is enhanced than when expressed with an unmodified pro sequence.

Table - CEA ELISA activity of A5B7 $F(ab')_2$-$(HCPB)_2$ expressed from COS cells

Protein expressed	Plasmid	Amount of A5B7 $F(ab')_2$-$(HCPB)_2$ (ng/ml)
none		0
$F(ab')_2$-$(HCPB)_2$	pMF53	0
+ pro	pMF53+pMF67	289
+ pro L	pMF53+pMF161	875
+ pro KKAA	pMF53+pMF165	615
+ pro SDYQRL	pMF53+pMF166	797

Example 10

Expression of [G251T,D253K]HCPB-His_6 -cMyc from COS cells by co-secretion of pro-L sequence

The procedure described in Reference Example 4 was repeated but with pMF48 (containing the gene for preHCPB) replaced by pMF133 (containing a gene for pre[G251T,D253K]HCPB-His_6 -cMyc), and pMF67 (containing the gene for the prepro sequence) was replaced by plasmid pMF161, containing a gene for a pro-L modified prepro sequence. Plasmid pMF133 was prepared as described in Reference Example 5.

Cell supernatants diluted 1 in 10 (250 µl) were analysed for enzyme activity against Hipp-Glu (5 h assay) as described in Example 14. The result is shown in Example 14 and is expressed as the percentage conversion of substrate into product (hippuric acid) at 37°.

Example 11

Expression of [G251T,D253K]HCPB-His_6 -cMyc from COS cells by co-secretion of pro-KDEL sequence

The procedure described in Example 10 was repeated but with pMF161, containing the gene for a pro-L modified prepro sequence, was replaced by plasmid pMF164 containing a gene for a pro-KDEL modified prepro sequence. The pro-KDEL sequence is the natural pro sequence with a C-terminal KDEL tetrapeptide residue. The amino acid sequence of pro-KDEL is shown in SEQ ID NO: 29.

Plasmid pMF164 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 30. The 365bp fragment was cloned into pBluescript to give pMF149 and subsequently into pEE12 (screening by PCR with SEQ ID NOS: 7 and 30) to give pMF164.

Cell supernatants diluted 1 in 10 (250 µl) were analysed for enzyme activity against Hipp-Glu (5h assay) as described in Example 14. The result is shown in Example 14 and is expressed as the percentage conversion of substrate into product (hippuric acid) at 37°.

Example 12

Expression of [G251T,D253K]HCPB-His_6 -cMyc from COS cells by co-secretion of pro-KKAA sequence

The procedure described in Example 10 was repeated but with pMF161, containing the gene for a pro-L modified prepro sequence, was replaced by plasmid pMF164 containing a gene for a pro-KKAA modified prepro sequence. The pro-KKAA sequence is the natural pro sequence with a C-terminal KKAA tetrapeptide residue. The amino acid sequence of pro-KKAA is shown in SEQ ID NO: 31

Plasmid pMF165 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 32. The 365bp fragment was cloned into pBluescript to give pMF145 and subsequently into pEE12 (Screening by PCR with SEQ ID NOS: 7 and 32) to give pMF165.

Cell supernatants diluted 1 in 10 (250 µl) were analysed for enzyme activity against Hipp-Glu (5 h assay) as described in Example 14. The result is shown in Example 14 and is expressed as the percentage conversion of substrate into product (hippuric acid) at 37°.

Example 13

Expression of [G251T,D253K]HCPB-His_6 -cMyc from COS cells by co-secretion of pro-SDYQRL sequence

The procedure described in Example 10 was repeated but with pMF161, containing the gene for a pro-L modified prepro sequence, was replaced by plasmid pMF164 containing a gene for a pro-SDYQRL modified prepro sequence. The pro-SDYQRL sequence is the natural pro sequence with a C-terminal SDYQRL hexapeptide residue. The amino acid sequence of pro-SDYQRL is shown in SEQ ID NO: 33.

Plasmid pMF166 was prepared by PCR from pMF18 as described for the unmodified prepro sequence, but using oligos SEQ ID NOS: 7 and 34. The 371bp fragment was cloned into pBluescript to give pMF148 and subsequently into pEE12 (screening by PCR with SEQ ID NOS: 7 and 34) to give pMF166.

Cell supernatants diluted 1 in 10 (250µl) were analysed for enzyme activity against Hipp-Glu (5 h assay) as described in Example 14. The result is shown in Example 14 and is expressed as the percentage conversion of substrate into product (hippuric acid) at 37°.

Example 14

Enzymic activity of recombinant [G251T,D253K]HCPB-His_6 -cMyc against Hipp-Glu.

COS cell supernatant (from Reference Example 5 and Examples 11 to 13) was assayed for its ability to convert hippuryl-L-glutamic acid (Hipp-Glu; described in Reference Example 1 of International Patent Application WO 97/07769, Zeneca Ltd, published 6 March, 1997) to hippuric acid using a HPLC assay.

The reaction mixture contained COS cell supernatant diluted 1 in 10 in 0.025 M Tris-HCL, pH 7.5 and 0.5 mM Hipp-Glu in 0.025 M Tris-HCL, pH 7.5 (total volume 250 µl). Samples were incubated for 5 h at 37°. The reactions were terminated by the addition of 250 µl of 30 % methanol, 70 % Phosphate Buffer, 0.2 % trifluoroacetic acid and the amount of hippuric acid generated was quantified by HPLC.

HPLC analysis was carried out using a Hewlett Packard 1090 Series 11 (with diode array) HPLC system. Samples (50 µl) were injected onto a reverse phase, base deactivated, octyl/octadecylsilane column (HICHROM Hi-RPB™) (25 cm) and separated

using a mobile phase of 15 % methanol, 85 % Phosphate Buffer at a flow rate of 1ml/min. The amount of product (hippuric acid, detected at 230nm) produced was determined from calibration curves generated with known amounts of hippuric acid (Sigma-H6375). The results are expressed as the percentage conversion of substrate into product in 5 h at 37°.

The data demonstrate that co-expression of [G251T,D253K]HCPB-His_6 -cMyc in the presence of pro sequences with additional C-terminal amino-acid residues is 2 to 9 fold higher than when expressed with an unmodified pro sequence.

Table - [G251T,D253K]HCPB-His_6 -cMyc expression from COS cells

		% hydrolysis of Hipp-Glu
Protein expressed	Plasmid	1:10 dilution
[G251T,D253K]HCPB-His_6 -cMyc	pMF133	0
+ pro	pMF133+pMF67	2.4
+ pro L	pMF133+pMF161	10.1
+ pro KDEL	pMF133+pMF164	12.9
+ pro KKAA	pMF133+pMF165	14.3
+ pro SDYQRL	pMF133+pMF166	21.2

Example 15

Expression of 806.077 $F(ab')_2$ -{[A248S,G251T,D253K]HCPB}$_2$ fusion protein from COS cells by co-secretion of pro-L sequence

This Example describes the preparation of a gene encoding a humanised Fd heavy chain fragment of antibody 806.077 linked to enzyme [A248S,G251T,D253K]HCPB and its co-expression with a gene encoding a humanised light chain of 806.077 and a gene encoding the pro-L modified prodomain of human carboxypeptidase B to give the $F(ab')_2$ protein with a molecule of [A248S,G251T,D253K]HCPB at the C-terminus of each of the heavy chain fragments. The constant and hinge regions of of the humanised Fd heavy chain fragment are derived from the human IgG3 antibody isotype. The expressed protein is also referred to as antibody-enzyme fusion protein.

(a) Preparation of a gene encoding humanised Fd heavy chain fragment of 806.077 linked to [A248S,G251T,D253K]HCPB and its cloning into pEE6'

A gene encoding humanised 806.077 Fd linked to [A248S,G251T,D253K]HCPB was generated by PCR from pZEN1921 (this plasmid is also named pMC60.3 and is described in Example 37 of International Patent Application WO 97/07769, Zeneca Ltd, published 6 March, 1997). A first PCR was set up with template pZEN1921 (2 ng) and oligonucleotides SEQ ID NO: 35 and SEQ ID NO: 36 (100 pM of each) in PCR Buffer (100 µl). The reaction was incubated at 94° for 5 min then thermostable DNA polymerase (2.5 u, 0.5 ml) was added and the mixture overlaid with mineral oil (100 µl) and the reaction mixture incubated at 94° for 1 min, 53° for 1 min and 72° for 2.5 min for 25 cycles, plus 10 min at 72°. The PCR product of 536 base pairs was isolated by electrophoresis on a 1 % agarose (Agarose type I, Sigma A-6013) gel followed by excision of the band from the gel and isolation of the DNA fragment.

A second PCR was set up with template IgG3-pBSIIKS+ (8.7ng, described in Reference Example 6) and oligonucleotides SEQ ID NO: 37 and SEQ ID NO: 38 and the 954 base pairs fragment isolated as described above. The products from the 2 PCRs were combined (either at 0.2, 1.0 or 5.0 ng/ml) in PCR Buffer as described above. The mixture was incubated for at 94° for 5 min then 10 cycles at 94° for 1 min and 63° for 4 min. Oligos SEQ ID NOS: 36 and 37 (100 pM of each) in PCR Buffer (50 µl) were added. After incubation at 94° for 3 min, the mixture was further incubated at 94° for 1.5 min, 53° for 2 min and 72° for 2 min for 25 cycles plus 10 min at 72°. In this process, the G base at position 508 in SEQ ID NO: 39 was changed to an A base.

The PCR product of 1434 base pairs was isolated by electrophoresis on a 1 % agarose gel, purified and digested with NheI (20 u) and XbaI (80 u) (New England Biolabs Inc.,) in a total volume of 100 µl containing 10 mM Tris HCl (pH7.9), 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT and BSA (100 µg/ml) for 4 h at 37°. The resulting fragment was again isolated by electrophoresis on a 1 % agarose gel and purified. In a similar digestion, vector pNG4-VHss-806.077huVH1-HuIgG2CH1' (10 µg; Example 11) was cut with NheI and XbaI then calf intestinal alkaline phosphatase (1µl; New England Biolabs, 10u/µl) was added to the digested plasmid to remove 5' phosphate groups and incubation continued at 37° for a further 30

minutes. Phosphatase activity was destroyed by incubation at 70^{O} for 10 minutes. The NheI-XbaI cut plasmid was purified from an agarose gel. The NheI-XbaI digested PCR product from above (about 500 ng) was ligated with the above cut plasmid DNA (about 200 ng) in 20 µl of a solution containing 50 mM Tris-Hcl (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 50 µg/ml BSA and 400u T4 DNA ligase at 25^{O} for 4h. A 1 µl aliquot of the reaction was used to transform 20 µl of competent E. coli DH5α cells. Transformed cells were plated onto L-agar plus 100 µg/ml ampicillin. Potential clones containing the gene for humanised 806.077 Fd-[A248S,G251T,D253K]HCPB were identified by PCR. Each clone was subjected to PCR as described above with oligonucleotides SEQ ID NOS: 40 and 41. A sample (10 µl) of the PCR reaction was analysed by electrophoresis on a 1 % agarose gel. Clones containing the required gene were identified by the presence of a 512 base pairs PCR product. Clones producing the 512 base pairs band were used for DNA minipreps. The DNA samples were checked by digestion with HindIII and XbaI for the presence of 3751 base pairs and 1862 base pairs fragments. Clones containing these fragments on digestion of the DNA with HindIII and XbaI were used for large scale plasmid DNA preparation and the sequence of the insert confirmed by DNA sequence analysis. The sequence of the expected insert is shown in SEQ ID NO: 42 Of the clones examined above, 2 contained the expected sequence but with a single base mutation. Clone 54 (also designated pMF195) had an T base at position 605 in SEQ ID NO: 42 in place of the A base, whereas clone 68 (also designated pMF198) had a C base at position 1825 instead of the expected T base. The sequence shown in SEQ ID NO: 42 was prepared from pMF195 and and pMF198 by digesting both (10 µg of each) with XmaI (10u) and XbaI (100 u) (New England Biolabs) in buffer (100 µl) containing 20 mM Tris acetate (pH 7.9) 50 mM potassium acetate, 10 mM Mg acetate, 1 mM DTT and BSA (100 µg/ml). The 215 base pairs fragment from pMF195 and the vector fragment from pMF198 (following treatment with alkaline phosphatase) were isolated from a 1 % agarose gel and ligated together as described previously. The ligation mix was used to transform competent DH5α cells. The transformed cells were plated onto L agar plus ampicillin and resulting colonies screened by digestion of the DNA with XmaI and XbaI for the presence of 5400 base pairs and 215 base pairs fragments. Positive clones were used for large scale plasmid DNA preparation and the sequence of the insert confirmed by DNA sequence analysis. The plasmid containing the 806.077 Fd-[A248S,G251T,D253K]HCPB gene from

clone number 102 was named pMF213. The HindIII-XbaI fragment from pMF213 was cloned into pEE6 [this is a derivative of pEE6.hCMV - Stephens and Cockett (1989) Nucleic Acids Research 17, 7110 - in which a HindIII site upstream of the hCMV promoter has been converted to a BglII site] in DH5α (screened by PCR with oligonucleotides SEQ ID NOS: 43 and 44 for a 2228 base pairs insert) to give pMF221.

(b) Preparation of a co-expression vector for expression of antibody-enzyme fusion protein

To generate vectors capable of expressing the antibody-enzyme fusion protein in eukaryotic cells, the GS-System™ (Celltech Biologics) was used (WO 87/04462, WO 89/01036, WO 86/05807 and WO 89/10404). The procedure requires cloning the humanised antibody light chain gene into the HindIII-XmaI region of vector pEE14. This vector is described by Bebbington in METHODS: A Companion to methods in Enzymology (1991) 2, 136-145. To construct the expression vector, plasmids pEE14 and pNG3-VKss-806.077HuVK4-HuCK-Neo (Reference Example 7) were digested with HinIII and XmaI as described above. The appropriate vector (from pEE14) and insert (732 base pairs from pNG3-VKss-806.077HuVK4-HuCK-Neo) from each digest were isolated from a 1 % agarose gel and ligated together and used to transform competent DH5α cells. The transformed cells were were plated onto L agar plus ampicillin (100 μg/ml). Colonies were screened by restriction analysis of isloated DNA for the presence of a 732 base pairs fragment on digestion of the DNA with HindIII and XmaI. Clones producing a 732 base pairs restriction fragment were used for large scale plasmid DNA preparation and the sequence of the insert confirmed by DNA sequence analysis. The plasmid containing the humanised light chain sequence of SEQ ID NO: 45 in pEE14 was named pEE14-806.077HuVK4-HuCK.

To make the co-expression vector, pMF221 (10 μg) was cut with BglII (20 u) and SalI (40 U) in buffer (100 μl) containing 10 mM Tris-HCl (pH 7.9), 150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT and BSA (100μg/ml) and the 4560 base pairs fragment isolated by agarose gel electrophoresis and purified. Similarly, pEE14-806.077HuVK4-HuCK was cut with BamHI (40 u) and SalI (40 u) and the 9.95 kb vector fragment isolated and ligated to the BglII-SalI fragment from pMF221 and cloned into DH5α. Colonies were screened by PCR with 2 sets of oligonucleotides (SEQ ID NOS: 40 and 41, and SEQ ID NOS: 46 and 47). Clones giving PCR products of 185 base pairs and 525 base pairs respectively

were characterised by DNA sequencing. A clone with the correct sequence was named pMF228 - light chain/Fd-mutant HCPB co-expression vector in DH5α. The humanised Fd-mutant HCPB sequence is shown in SEQ ID NO: 48. Residues 1 to 19 are the signal sequence, residues 20 to 242 are the humanised variable and IgG3 CH1 region, residues 243 to 306 are the IgG3 hinge region and residues 307 to 613 are the mutant HCPB sequence with the changes at residues 248, 251 and 253 from the human HCPB sequence. The changes in the HCPB sequence occur in SEQ ID NO: 48 at postions 554 (Ser), 557 (Thr) and 559 (Lys) respectively.

(c) Preparation of a vector for expression of the prodomain of proHCPB

A second eukaryotic expression plasmid, pMF161, containing a gene for the prepro sequence, for secretion of the prodomain with an additional C-terminal leucine residue (termed pro-L), of preproHCPB was prepared as described in Example 1.

(d) Expression of antibody-enzyme fusion protein in eukaryotic cells

For expression in eukaryotic cells, vectors containing genes capable of expressing the antibody enzyme-fusion protein (pMF228) and the pro-L sequence (pMF161) were co-transfected into COS-7 cells as described in Reference Example 4. Cell supernatants diluted 1 in 10 (250 µl) were analysed for enzyme activity against Hipp-Glu (5 h assay) as described in Example 19. The result is shown in Example 19 and is expressed as the percentage conversion of substrate into product at 37°.

(e) Western analysis

Western blot analysis was performed as described as follows. Aliquots (20 µl) of each supernatant sample were mixed with an equal volume of sample buffer (62.5 mM Tris, pH 6.8, 1 % SDS, 10 % sucrose and 0.05 % bromophenol blue) with and without reductant. The samples were incubated at 65° for 10 minutes before electrophoresis on a 8-18 % acrylamide gradient gel (EXCEL™ gel system from Pharmacia Biotechnology Products) in a MULTIPHOR™ II apparatus (LKB Produkter AB) according to the manufacturer's instructions. After electrophoresis, the separated proteins were transfered to a membrane (HYBOND™ C-Super,Amersham International) using a NOVABLOT™ apparatus (LKB Produkter AB) according to protocols provided by the manufacturer. After blotting, the membrane was air dried.

The presence of antibody fragments was detected by the use of an anti-human kappa antibody (Sigma A7164, goat anti-human Kappa light chain peroxidase conjugate) used at 1:2500 dilution. The presence of human antibody fragments was visualised using a chemiluminescence system (ECL™ detection system, Amersham International).

Example 16

Expression of 806.077 F(ab')$_2$ -{[A248S,G251T,D253K]HCPB}$_2$ fusion protein from COS cells by co-secretion of pro-KDEL sequence

The procedure described in Example 15 was repeated but with pMF161, containing the gene for a pro-L modified prepro sequence, was replaced by plasmid pMF164 containing a gene for a pro-KDEL modified prepro sequence. The pro-KDEL sequence is the natural pro sequence with a C-terminal KDEL tetrapeptide residue. The amino acid sequence of pro-KDEL is shown in SEQ ID NO: 29. Plasmid pMF165 is described in Example 2.

Cell supernatants diluted 1 in 10 (250µl) were analysed for enzyme activity against Hipp-Glu (5 h assay) as described in Example 19. The result is shown in Example 19 and is expressed as the percentage conversion of substrate into product (hippuric acid) at 37°.

Example 17

Expression of 806.077 F(ab')$_2$ -{[A248S,G251T,D253K]HCPB}$_2$ fusion protein from COS cells by co-secretion of pro-KKAA sequence

The procedure described in Example 15 was repeated but with pMF161, containing the gene for a pro-L modified prepro sequence, was replaced by plasmid pMF165 containing a gene for a pro-KKAA modified prepro sequence. The pro-KKAA sequence is the natural pro sequence with a C-terminal KKAA tetrapeptide residue. The amino acid sequence of pro-KKAA is shown in SEQ ID NO: 31. Plasmid pMF165 is described in Example 3.

Cell supernatants diluted 1 in 10 (250 µl) were analysed for enzyme activity against Hipp-Glu (5 h assay) as described in Example 19. The result is shown in Example 19 and is expressed as the percentage conversion of substrate into product (hippuric acid) at 37°.

Example 18

Expression of 806.077 F(ab')$_2$ -{[A248S,G251T,D253K]HCPB}$_2$ fusion protein from COS cells by co-secretion of pro-SDYQRL sequence

The procedure described in Example 15 was repeated but with pMF161, containing the gene for a pro-L modified prepro sequence, was replaced by plasmid pMF166 containing a gene for a pro- SDYQRL modified prepro sequence. The pro- SDYQRL sequence is the natural pro sequence with a C-terminal KKAA tetrapeptide residue. The amino acid sequence of pro- SDYQRL is shown in SEQ ID NO: 33. Plasmid pMF166 is described in Example 4.

Cell supernatants diluted 1 in 10 (250 µl) were analysed for enzyme activity against Hipp-Glu (5 h assay) as described in Example 19. The result is shown in Example 19 and is expressed as the percentage conversion of substrate into product (hippuric acid) at 37°.

Example 19

Enzymic activity and antigen binding activity of recombinant 806.077 F(ab')$_2$ - {[A248S,G251T,D253K]HCPB}$_2$ fusion protein

COS cell supernatant (from Examples 15 to 18) was assayed for its ability to convert hippuryl-L-glutamic acid (Hipp-Glu; described in Reference Example 1 of International Patent Application WO 97/07769, Zeneca Ltd, published 6 March, 1997) to hippuric acid using a HPLC assay. As a control sample, the procedure described in Example 15 was repeated but with the plasmid pMF161 (containing the gene for a pro-L modified prepro sequence) replaced by pMF67 (containing the unmodified prepro sequence) as described in Reference Example 4.

The reaction mixture (250 µl) contained COS cell supernatant diluted 1 in 5 in 0.025 M Tris-HCL, pH 7.5 (125 µl), and 0.5 mM Hipp-Glu in 0.025 M Tris-HCL, pH 7.5. Samples were incubated for 5 h at 37°. The reactions were terminated by the addition of 250 µl of 30 % methanol, 70 % Phosphate Buffer , 0.2 % trifluoroacetic acid and the amount of hippuric acid generated was quantified by HPLC.

HPLC analysis was carried out using a Hewlett Packard 1090 Series 11 (with diode array) HPLC system. Samples (50 µl) were injected onto a HICHROM Hi-RPB™ column (25 cm) and separated using a mobile phase of 15 % methanol, 85 % Phosphate Buffer at a flow rate of 1 ml/min. The amount of product (hippuric acid, detected at 230 nm) produced

was determined from calibration curves generated with known amounts of hippuric acid (Sigma-H6375). The results are expressed as the percentage conversion of substrate into product in 5 h at 37°.

The data demonstrate that co-expression of 806.077 $F(ab')_2$ - $\{[A248S,G251T,D253K]HCPB\}_2$ fusion protein in the presence of pro sequences with additional C-terminal amino-acid residues is higher than when expressed with an unmodified pro sequence.

Table - Enzyme activity of 806.077 $F(ab')_2$ -$\{[A248S,G251T,D253K]HCPB\}_2$ fusion protein expression from COS cells

		% hydrolysis of Hipp-Arg
Protein expressed	Plasmid	1:10 dilution
none		0
806.077 $F(ab')_2$ - $\{[A248S,G251T, D253K]HCPB\}_2$	pMF228	0
+ pro	pMF228+pMF67	7.3
+ pro L	pMF228+pMF161	22.4
+pro KDEL	pMF228+pMF164	9.5
+ pro KKAA	pMF228+pMF165	14.8
+ pro SDYQRL	pMF228+pMF166	12.3

COS cell supernatant were also assayed for antigen binding activity in a CEA ELISA assay performed essentially as described in Reference Example 13 of International Patent Application WO 97/07769, Zeneca Ltd, published 6 March, 1997, but with 250 ng/well used instead of 50 ng/well of CEA. The results are expressed as concentrations of fusion protein present in the supernatant. The amount of of 806.077 $F(ab')_2$ - $\{[A248S,G251T,D253K]HCPB\}_2$ fusion protein present in the COS cell supernatants was determined from calibration curves generated with known amounts of of 806.077 $F(ab')_2$ - $\{[A248S,G251T,D253K]HCPB\}_2$ fusion protein.

The data demonstrate that co-expression of of 806.077 F(ab')$_2$ - {[A248S,G251T,D253K]HCPB}$_2$ fusion protein in the presence of pro sequences with additional C-terminal amino-acid residues is enhanced compared with expression in the presence of an unmodified pro sequence.

Table - CEA ELISA activity of 806.077 F(ab')$_2$ -{[A248S,G251T,D253K]HCPB}$_2$ expressed from COS cells

Protein expressed	Plasmid	Amount of fusion protein (ng/ml)
none		0
806.077 F(ab')$_2$ - {[A248S,G251T,D253K]HCPB}$_2$ fusion protein	pMF228	200
+ pro	pMF228+pMF67	1410
+ pro L	pMF228+pMF161	6530
+pro KDEL	PMF228+pMF164	2460
+ pro KKAA	pMF228+pMF165	4510
+ pro SDYQRL	pMF228+pMF166	4140

ES70217

AFG/MB : 08JAN98

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Zeneca Limited
(B) STREET: 15 Stanhope Gate
(C) CITY: London
(D) STATE: England
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): W1Y 6LN
(G) TELEPHONE: 0171 304 5000
(H) TELEFAX: 0171 304 5151
(I) TELEX: 0171 304 2042

(ii) TITLE OF INVENTION: PROTEINS

(iii) NUMBER OF SEQUENCES: 53

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GTTGGAGCTC TTGGTTCTGG 20

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CAAGGCCTCG AGCTTTCTCA AC 22

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GTTTGATTCT AGAGTTCGTG C 21

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TTGTAAAACG ACGGCCAGTG AG 22

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GAAACAGCTA TGACCATGAT TACG 24

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CAGACTCTGC AGCAGGTCCA CAG 23

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGACCTGCTG CAGAGTCTG 19

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCCTGTGCTC AATATTGATG G 21

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CCGTGTTAAA GCAGAAGATA CTG 23

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GCTACTGTGA AAGAACTTGC CTC 23

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1263 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GAGCTCTTGG TTCTGGTGAC TGTGGCCCTG GCATCTGCTC ATCATGGTGG TGAGCACTTT 60

GAAGGCGAGA AGGTGTTCCG TGTTAACGTT GAAGATGAAA ATCACATTAA CATAATCCGC 120

GAGTTGGCCA GCACGACCCA GATTGACTTC TGGAAGCCAG ATTCTGTCAC ACAAATCAAA 180

CCTCACAGTA CAGTTGACTT CCGTGTTAAA GCAGAAGATA CTGTCACTGT GGAGAATGTT 240

CTAAAGCAGA ATGAACTACA ATACAAGGTA CTGATAAGCA ACCTGAGAAA TGTGGTGGAG 300

GCTCAGTTTG ATAGCCGGGT TCGTGCAACA GGACACAGTT ATGAGAAGTA CAACAAGTGG 360

GAAACGATAG AGGCTTGGAC TCAACAAGTC GCCACTGAGA ATCCAGCCCT CATCTCTCGC 420

AGTGTTATCG GAACCACATT TGAGGGACGC GCTATTTACC TCCTGAAGGT TGGCAAAGCT 480

GGACAAAATA AGCCTGCCAT TTTCATGGAC TGTGGTTTCC ATGCCAGAGA GTGGATTTCT 540

CCTGCATTCT GCCAGTGGTT TGTAAGAGAG GCTGTTCGTA CCTATGGACG TGAGATCCAA 600

GTGACAGAGC TTCTCGACAA GTTAGACTTT TATGTCCTGC CTGTGCTCAA TATTGATGGC 660

TACATCTACA CCTGGACCAA GAGCCGATTT TGGAGAAAGA CTCGCTCCAC CCATACTGGA 720

TCTAGCTGCA TTGGCACAGA CCCCAACAGA AATTTTGATG CTGGTTGGTG TGAAATTGGA 780

GCCTCTCGAA ACCCCTGTGA TGAAACTTAC TGTGGACCTG CCGCAGAGTC TGAAAAGGAA 840

ACCAAGGCCC TGGCTGATTT CATCCGCAAC AAACTCTCTT CCATCAAGGC ATATCTGACA 900

ATCCACTCGT ACTCCCAAAT GATGATCTAC CCTTACTCAT ATGCTTACAA ACTCGGTGAG 960

AACAATGCTG AGTTGAATGC CCTGGCTAAA GCTACTGTGA AAGAACTTGC CTCACTGCAC 1020

GGCACCAAGT ACACATATGG CCCGGGAGCT ACAACAATCT ATCCTGCTGC TGGGGGCTCT 1080

GACGACTGGG CTTATGACCA AGGAATCAGA TATTCCTTCA CCTTTGAACT TCGAGATACA 1140

GGCAGATATG GCTTTCTCCT TCCAGAATCC CAGATCCGGG CTACCTGCGA GGAGACCTTC 1200

CTGGCAATCA AGTATGTTGC CAGCTACGTC CTGGAACACC TGTACTAGTT GAGAAAGCTC 1260

GAG 1263

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 415 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
Glu Leu Leu Val Leu Val Thr Val Ala Leu Ala Ser Ala His His Gly
1               5                  10                  15

Gly Glu His Phe Glu Gly Glu Lys Val Phe Arg Val Asn Val Glu Asp
            20                  25                  30

Glu Asn His Ile Asn Ile Ile Arg Glu Leu Ala Ser Thr Thr Gln Ile
        35                  40                  45

Asp Phe Trp Lys Pro Asp Ser Val Thr Gln Ile Lys Pro His Ser Thr
    50                  55                  60

Val Asp Phe Arg Val Lys Ala Glu Asp Thr Val Thr Val Glu Asn Val
65                  70                  75                  80

Leu Lys Gln Asn Glu Leu Gln Tyr Lys Val Leu Ile Ser Asn Leu Arg
                85                  90                  95

Asn Val Val Glu Ala Gln Phe Asp Ser Arg Val Arg Ala Thr Gly His
            100                 105                 110

Ser Tyr Glu Lys Tyr Asn Lys Trp Glu Thr Ile Glu Ala Trp Thr Gln
        115                 120                 125

Gln Val Ala Thr Glu Asn Pro Ala Leu Ile Ser Arg Ser Val Ile Gly
    130                 135                 140

Thr Thr Phe Glu Gly Arg Ala Ile Tyr Leu Leu Lys Val Gly Lys Ala
145                 150                 155                 160

Gly Gln Asn Lys Pro Ala Ile Phe Met Asp Cys Gly Phe His Ala Arg
                165                 170                 175

Glu Trp Ile Ser Pro Ala Phe Cys Gln Trp Phe Val Arg Glu Ala Val
            180                 185                 190

Arg Thr Tyr Gly Arg Glu Ile Gln Val Thr Glu Leu Leu Asp Lys Leu
        195                 200                 205

Asp Phe Tyr Val Leu Pro Val Leu Asn Ile Asp Gly Tyr Ile Tyr Thr
    210                 215                 220

Trp Thr Lys Ser Arg Phe Trp Arg Lys Thr Arg Ser Thr His Thr Gly
225                 230                 235                 240
```

```
Ser Ser Cys Ile Gly Thr Asp Pro Asn Arg Asn Phe Asp Ala Gly Trp
                245                 250                 255

Cys Glu Ile Gly Ala Ser Arg Asn Pro Cys Asp Glu Thr Tyr Cys Gly
            260                 265                 270

Pro Ala Ala Glu Ser Glu Lys Glu Thr Lys Ala Leu Ala Asp Phe Ile
        275                 280                 285

Arg Asn Lys Leu Ser Ser Ile Lys Ala Tyr Leu Thr Ile His Ser Tyr
    290                 295                 300

Ser Gln Met Met Ile Tyr Pro Tyr Ser Tyr Ala Tyr Lys Leu Gly Glu
305                 310                 315                 320

Asn Asn Ala Glu Leu Asn Ala Leu Ala Lys Ala Thr Val Lys Glu Leu
                325                 330                 335

Ala Ser Leu His Gly Thr Lys Tyr Thr Tyr Gly Pro Gly Ala Thr Thr
            340                 345                 350

Ile Tyr Pro Ala Ala Gly Gly Ser Asp Asp Trp Ala Tyr Asp Gln Gly
        355                 360                 365

Ile Arg Tyr Ser Phe Thr Phe Glu Leu Arg Asp Thr Gly Arg Tyr Gly
    370                 375                 380

Phe Leu Leu Pro Glu Ser Gln Ile Arg Ala Thr Cys Glu Glu Thr Phe
385                 390                 395                 400

Leu Ala Ile Lys Tyr Val Ala Ser Tyr Val Leu Glu His Leu Tyr
                405                 410                 415
```

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GCCGGGTTTG CGCAACTGGT CACTCTTACG AGAAG 35

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 88 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CCGGAATTCT TATTAGTTCA GGTCCTCCTC AGAGATCAGC TTCTGCTCCT CGAACTCATG 60

GTGGTGATGG TGGTGGTACA GGTGTTCC 88

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TTAGCGGATC CTGCCTGACG GT 22

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGCTGGATTC TCAGTGGCGA CTT 23

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ACCTCTAGGG TCCCCAATTA 20

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CAAGTCGCCA CTGAGAATCC AGC 23

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CTCTAGGAAT TCTTATTAGT ACAGGTGTTC CAGGACGTAG C 41

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 108 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CCCAAGCTTG CCGCCACCAT GTTGGCAGTC TTGGTTCTGG TGACTGTGGC CCTGGCATCT 60

GCTGCAACAG GACACAGTTA TGAGAAGTAC AACAAGTGGG AAACGATA 108

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

AACAGCTATG ACCATG 16

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GTAAAACGAC GGCCAGT 17

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

TCGCTATTAC CATGGTGATG CGGTTTTGGC 30

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:
CAGACTCTGC AGCAGGTCCA CAG 23

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

CCCAAGCTTG CCGCCACCAT GTTGGCACTC TTGGTTCTGG TGACTGTGGC CCTG 54

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CTCATAACTG AATTCTTATT AACGAACCCG GCTATCAAA 39

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

His His Gly Gly Glu His Phe Glu Gly Glu Lys Val Phe Arg Val Asn
1 5 10 15

Val Glu Asp Glu Asn His Ile Asn Ile Ile Arg Glu Leu Ala Ser Thr
20 25 30

Thr Gln Ile Asp Phe Trp Lys Pro Asp Ser Val Thr Gln Ile Lys Pro
35 40 45

His Ser Thr Val Asp Phe Arg Val Lys Ala Glu Asp Thr Val Thr Val
50 55 60

Glu Asn Val Leu Lys Gln Asn Glu Leu Gln Tyr Lys Val Leu Ile Ser
65 70 75 80

Asn Leu Arg Asn Val Val Glu Ala Gln Phe Asp Ser Arg Val Arg Leu
85 90 95

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GGCTGCAGGA ATTCTTATTA TAGACGAACC CGGCTATCAA ACTGAGC 47

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

His His Gly Gly Glu His Phe Glu Gly Glu Lys Val Phe Arg Val Asn
1 5 10 15

Val Glu Asp Glu Asn His Ile Asn Ile Ile Arg Glu Leu Ala Ser Thr
20 25 30

Thr Gln Ile Asp Phe Trp Lys Pro Asp Ser Val Thr Gln Ile Lys Pro
35 40 45

His Ser Thr Val Asp Phe Arg Val Lys Ala Glu Asp Thr Val Thr Val
50 55 60

Glu Asn Val Leu Lys Gln Asn Glu Leu Gln Tyr Lys Val Leu Ile Ser
65 70 75 80

Asn Leu Arg Asn Val Val Glu Ala Gln Phe Asp Ser Arg Val Arg Lys
85 90 95

Asn Glu Leu

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GGCTGCAGGA ATTCTTATTA TAGCTCATCC TTACGAACCC GGCTATCAAA CTGAGC 56

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

His His Gly Gly Glu His Phe Glu Gly Glu Lys Val Phe Arg Val Asn
1 5 10 15

Val Glu Asp Glu Asn His Ile Asn Ile Ile Arg Glu Leu Ala Ser Thr
20 25 30

Thr Gln Ile Asp Phe Trp Lys Pro Asp Ser Val Thr Gln Ile Lys Pro
35 40 45

His Ser Thr Val Asp Phe Arg Val Lys Ala Glu Asp Thr Val Thr Val
50 55 60

Glu Asn Val Leu Lys Gln Asn Glu Leu Gln Tyr Lys Val Leu Ile Ser
65 70 75 80

Asn Leu Arg Asn Val Val Glu Ala Gln Phe Asp Ser Arg Val Arg Lys
85 90 95

Lys Ala Ala

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GGCTGCAGGA ATTCTTATTA GGCTGCCTTC TTACGAACCC GGCTATCAAA CTGAGC 56

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 101 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

His His Gly Gly Glu His Phe Glu Gly Glu Lys Val Phe Arg Val Asn
1 5 10 15

Val Glu Asp Glu Asn His Ile Asn Ile Ile Arg Glu Leu Ala Ser Thr
20 25 30

Thr Gln Ile Asp Phe Trp Lys Pro Asp Ser Val Thr Gln Ile Lys Pro
35 40 45

His Ser Thr Val Asp Phe Arg Val Lys Ala Glu Asp Thr Val Thr Val
50 55 60

Glu Asn Val Leu Lys Gln Asn Glu Leu Gln Tyr Lys Val Leu Ile Ser
65 70 75 80

Asn Leu Arg Asn Val Val Glu Ala Gln Phe Asp Ser Arg Val Arg Ser
85 90 95

Asn Tyr Gln Arg Leu
100

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 62 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GGCTGCAGGA ATTCTTATTA TAGACGCTGG TAATCGCTAC GAACCCGGCT ATCAAACTGA 60

GC 62

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCCAGCACCT GAACTCCTGG GAGGAGCAAC AGGACACAGT TATGAGAAGT ACAA 54

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GGGGGTCTAG ATTATTAGTA CAGGTGTTCC AGGACGTAGC TGGCAACATA 50

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GGGGGAGCTC GGCTAGCACC AAGGGCCCAT CGGTCTTCCC CCTGGC 46

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 55 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

TTGTACTTCT CATAACTGTG TCCTGTTGCT CCTCCCAGGA GTTCAGGTGC TGGGC 55

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 520 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

GAGCTCGGCT AGCACCAAGG GCCCATCGGT CTTCCCCCTG GCGCCCTGCT CCAGGAGCAC 60

CTCTGGGGGC ACAGCGGCCC TGGGCTGCCT GGTCAAGGAC TACTTCCCCG AACCGGTGAC 120

GGTGTCGTGG AACTCAGGCG CCCTGACCAG CGGCGTGCAC ACCTTCCCGG CTGTCCTACA 180

GTCCTCAGGA CTCTACTCCC TCAGCAGCGT GGTGACCGTG CCCTCCAGCA GCTTGGGCAC 240

CCAGACCTAC ACCTGCAACG TGAATCACAA GCCCAGCAAC ACCAAGGTGG ACAAGAGAGT 300

GGAGCTGAAA ACCCCACTCG GTGACACAAC TCACACGTGC CCTAGGTGTC CTGAACCTAA 360

ATCTTGTGAC ACACCTCCCC CGTGCCCACG GTGCCCAGAG CCCAAATCTT GCGACACGCC 420

CCCACCGTGT CCCAGATGTC CTGAACCAAA GAGCTGTGAC ACTCCACCGC CCTGCCCGAG 480

GTGCCCAGCA CCTGAACTCC TGGGAGGGTA ATAGCCCGGG 520

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

GCCTGTGCTC AATATTGATG G 21

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GGAGAAAGCC ATATCTGCCT G 21

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1870 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

AAGCTTGCCG CCACCATGAA GTTGTGGCTG AACTGGATTT TCCTTGTAAC ACTTTTAAAT 60

GGAATTCAGT GTGAGGTGCA GCTGCAGCAG AGCGGTCCAG GTCTCGTACG GCCTAGCCAG 120

ACCCTGAGCC TCACGTGCAC CGCATCTGGC TTCAACATTA AGGACAATTA CATGCACTGG 180

GTGAGACAGC CACCTGGACG AGGCCTTGAG TGGATTGGAT GGATTGACCC TGAGAATGGT 240

GACACTGAGT ACGCACCTAA GTTTCGCGGC CGCGTGACAA TGCTGGCAGA CACTAGTAAG 300

AACCAGTTCA GCCTGAGACT CAGCAGCGTG ACAGCCGCCG ACACCGCGGT CTATTATTGT 360

CACGTCCTGA TATACGCCGG GTATCTGGCA ATGGACTACT GGGGCCAAGG GACCCTCGTC 420

ACCGTGAGCT CGGCTAGCAC CAAGGGCCCA TCGGTCTTCC CCCTGGCGCC CTGCTCCAGG 480

AGCACCTCTG GGGGCACAGC GGCCCTGGGC TGCCTGGTCA AGGACTACTT CCCCGAACCG 540

GTGACGGTGT CGTGGAACTC AGGCGCCCTG ACCAGCGGCG TGCACACCTT CCCGGCTGTC 600

CTACAGTCCT CAGGACTCTA CTCCCTCAGC AGCGTGGTGA CCGTGCCCTC CAGCAGCTTG 660

GGCACCCAGA CCTACACCTG CAACGTGAAT CACAAGCCCA GCAACACCAA GGTGGACAAG 720

AGAGTGGAGC TGAAAACCCC ACTCGGTGAC ACAACTCACA CGTGCCCTAG GTGTCCTGAA 780

CCTAAATCTT GTGACACACC TCCCCCGTGC CCACGGTGCC CAGAGCCCAA ATCTTGCGAC 840

ACGCCCCCAC CGTGTCCCAG ATGTCCTGAA CCAAAGAGCT GTGACACTCC ACCGCCCTGC 900

CCGAGGTGCC CAGCACCTGA ACTCCTGGGA GGAGCAACAG GACACAGTTA TGAGAAGTAC 960

AACAAGTGGG AAACGATAGA GGCTTGGACT CAACAAGTCG CCACTGAGAA TCCAGCCCTC 1020

ATCTCTCGCA GTGTTATCGG AACCACATTT GAGGGACGCG CTATTTACCT CCTGAAGGTT 1080

GGCAAAGCTG GACAAAATAA GCCTGCCATT TTCATGGACT GTGGTTTCCA TGCCAGAGAG 1140

TGGATTTCTC CTGCATTCTG CCAGTGGTTT GTAAGAGAGG CTGTTCGTAC CTATGGACGT 1200

GAGATCCAAG TGACAGAGCT TCTCGACAAG TTAGACTTTT ATGTCCTGCC TGTGCTCAAT 1260

ATTGATGGCT ACATCTACAC CTGGACCAAG AGCCGATTTT GGAGAAAGAC TCGCTCCACC 1320

CATACTGGAT CTAGCTGCAT TGGCACAGAC CCCAACAGAA ATTTTGATGC TGGTTGGTGT 1380

GAAATTGGAG CCTCTCGAAA CCCCTGTGAT GAAACTTACT GTGGACCTGC CGCAGAGTCT 1440

GAAAAGGAGA CCAAGGCCCT GGCTGATTTC ATCCGCAACA AACTCTCTTC CATCAAGGCA 1500

TATCTGACAA TCCACTCGTA CTCCCAAATG ATGATCTACC CTTACTCATA TGCTTACAAA 1560

CTCGGTGAGA ACAATGCTGA GTTGAATGCC CTGGCTAAAG CTACTGTGAA AGAACTTGCC 1620

TCACTGCACG GCACCAAGTA CACATATGGC CCGGGAGCTA CAACAATCTA TCCTTCTGCT 1680

GGGACTTCTA AAGACTGGGC TTATGACCAA GGAATCAGAT ATTCCTTCAC CTTTGAACTT 1740

CGAGATACAG GCAGATATGG CTTTCTCCTT CCAGAATCCC AGATCCGGGC TACCTGCGAG 1800

GAGACCTTCC TGGCAATCAA GTATGTTGCC AGCTACGTCC TGGAACACCT GTACTAATAA 1860

TCTAGAGAGA 1870

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TCGCTATTAC CATGGTGATG CGGTTTTGGC 30

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGCTGGATTC TCAGTGGCGA CTT 23

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 321 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GACATCCAGA TGACCCAGAG CCCAAGCAGC CTGAGCGCTA GCGTGGGTGA CAGAGTGACC 60

ATCACGTGTA GTGCCAGCTC AAGTGTAACT TACATGCACT GGTACCAGCA GAAGCCAGGT 120

AAGGCTCCAA AGCTGTGGAT CTACAGCACA TCCAACCTGG CTTCTGGTGT GCCAAGCAGA 180

TTCTCCGGAA GCGGTAGCGG CACCGACTAC ACCTTCACCA TCAGCAGCCT CCAGCCAGAG 240

GATATCGCCA CCTACTACTG CCAGCAGAGG AGTACTTACC CGCTCACGTT CGGCCAAGGG 300

ACCAAGCTCG AGATCAAACG G 321

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CACAACAGAG GCAGTTCC 18

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CACCTTCACC ATCAGCAGCC 20

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 613 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

Met Lys Leu Trp Leu Asn Trp Ile Phe Leu Val Thr Leu Leu Asn Gly
1 5 10 15

Ile Gln Cys Glu Val Gln Leu Gln Gln Ser Gly Pro Gly Leu Val Arg
20 25 30

Pro Ser Gln Thr Leu Ser Leu Thr Cys Thr Ala Ser Gly Phe Asn Ile
35 40 45

```
Lys Asp Asn Tyr Met His Trp Val Arg Gln Pro Pro Gly Arg Gly Leu
    50                  55                  60

Glu Trp Ile Gly Trp Ile Asp Pro Glu Asn Gly Asp Thr Glu Tyr Ala
65                  70                  75                  80

Pro Lys Phe Arg Gly Arg Val Thr Met Leu Ala Asp Thr Ser Lys Asn
                85                  90                  95

Gln Phe Ser Leu Arg Leu Ser Ser Val Thr Ala Ala Asp Thr Ala Val
            100                 105                 110

Tyr Tyr Cys His Val Leu Ile Tyr Ala Gly Tyr Leu Ala Met Asp Tyr
        115                 120                 125

Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly
    130                 135                 140

Pro Ser Val Phe Pro Leu Ala Pro Cys Ser Arg Ser Thr Ser Gly Gly
145                 150                 155                 160

Thr Ala Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val
                165                 170                 175

Thr Val Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe
            180                 185                 190

Pro Ala Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val
        195                 200                 205

Thr Val Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Thr Cys Asn Val
    210                 215                 220

Asn His Lys Pro Ser Asn Thr Lys Val Asp Lys Arg Val Glu Leu Lys
225                 230                 235                 240

Thr Pro Leu Gly Asp Thr Thr His Thr Cys Pro Arg Cys Pro Glu Pro
                245                 250                 255

Lys Ser Cys Asp Thr Pro Pro Pro Cys Pro Arg Cys Pro Glu Pro Lys
            260                 265                 270

Ser Cys Asp Thr Pro Pro Pro Cys Pro Arg Cys Pro Glu Pro Lys Ser
        275                 280                 285
```

```
Cys Asp Thr Pro Pro Pro Cys Pro Arg Cys Pro Ala Pro Glu Leu Leu
    290                 295                 300

Gly Gly Ala Thr Gly His Ser Tyr Glu Lys Tyr Asn Lys Trp Glu Thr
305                 310                 315                 320

Ile Glu Ala Trp Thr Gln Gln Val Ala Thr Glu Asn Pro Ala Leu Ile
                325                 330                 335

Ser Arg Ser Val Ile Gly Thr Thr Phe Glu Gly Arg Ala Ile Tyr Leu
            340                 345                 350

Leu Lys Val Gly Lys Ala Gly Gln Asn Lys Pro Ala Ile Phe Met Asp
        355                 360                 365

Cys Gly Phe His Ala Arg Glu Trp Ile Ser Pro Ala Phe Cys Gln Trp
    370                 375                 380

Phe Val Arg Glu Ala Val Arg Thr Tyr Gly Arg Glu Ile Gln Val Thr
385                 390                 395                 400

Glu Leu Leu Asp Lys Leu Asp Phe Tyr Val Leu Pro Val Leu Asn Ile
                405                 410                 415

Asp Gly Tyr Ile Tyr Thr Trp Thr Lys Ser Arg Phe Trp Arg Lys Thr
            420                 425                 430

Arg Ser Thr His Thr Gly Ser Ser Cys Ile Gly Thr Asp Pro Asn Arg
        435                 440                 445

Asn Phe Asp Ala Gly Trp Cys Glu Ile Gly Ala Ser Arg Asn Pro Cys
    450                 455                 460

Asp Glu Thr Tyr Cys Gly Pro Ala Ala Glu Ser Glu Lys Glu Thr Lys
465                 470                 475                 480

Ala Leu Ala Asp Phe Ile Arg Asn Lys Leu Ser Ser Ile Lys Ala Tyr
                485                 490                 495

Leu Thr Ile His Ser Tyr Ser Gln Met Met Ile Tyr Pro Tyr Ser Tyr
            500                 505                 510

Ala Tyr Lys Leu Gly Glu Asn Asn Ala Glu Leu Asn Ala Leu Ala Lys
        515                 520                 525
```

Ala Thr Val Lys Glu Leu Ala Ser Leu His Gly Thr Lys Tyr Thr Tyr
530 535 540

Gly Pro Gly Ala Thr Thr Ile Tyr Pro Ser Ala Gly Thr Ser Lys Asp
545 550 555 560

Trp Ala Tyr Asp Gln Gly Ile Arg Tyr Ser Phe Thr Phe Glu Leu Arg
565 570 575

Asp Thr Gly Arg Tyr Gly Phe Leu Leu Pro Glu Ser Gln Ile Arg Ala
580 585 590

Thr Cys Glu Glu Thr Phe Leu Ala Ile Lys Tyr Val Ala Ser Tyr Val
595 600 605

Leu Glu His Leu Tyr
610

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AAGCTTTCCC GCGGCGACAT CCAGATGACC CAGAGCCCAA GCAGCCTGAG CGCTAGCGTG 60

GGTGACAGAG TGACCATCAC GTGTAGTGCC AGCTCAAGTG TAACTTACAT GCACTGGTAC 120

CAGCAGAAGC CAGGTAAGGC TCCAAAGCTG CTGATCTACA GCACATCCAA CCTGGCTTCT 180

GGTGTGCCAA GCAGATTCTC CGGAAGCGGT AGCGGCACCG ACTACACCTT CACCATCAGC 240

AGCCTCCAGC CAGAGGATAT CGCCACCTAC TACTGCCAGC AGAGGAGTAC TTACCCGCTC 300

ACGTTCGGCC AAGGGACCAA GCTCGAGATC AAACGGACTA GT 342

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CGTATTAGTC ATCGCTATTA CC 22

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

GTTGGATGTG CTGTAGATCC ACAGCTTTGG AGCCTTACC 39

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

TCCGTTTGAT CTCGAGCTTG G 21

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

GGTAAGGCTC CAAAGCTGTG GATCTACAGC ACATCCAAC 39

ES70217

AFG/MB : 08JAN98

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 16, line 11-13

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

National Collection of Industrial and Marine Bacteria

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen
Scotland AB2 1RY

Date of deposit	Accession Number
11-Oct-93	NCIMB 40589

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

"In respect of all designated states in which such action is possible and to the extent that it is legally permissible under the law of the designated state, it is requested that a sample of the deposited micro-organism(s) be made available only by the issue thereof to an independent expert, in accordance with the relevant patent legislation, *e.g.* EPC Rule 28(4), U.K. Rule 17(3), Australian Regulation 3.25(3) and generally similar provisions *mutatis mutandis* for any other designated state".

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 6, line 9-12

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

European Collection of Animal Cell Cultures

Address of depositary institution *(including postal code and country)*

PHLS Centre for Applied Microbiology & Research
Porton Down
Salisbury
Wiltshire SP4 0JG
Great Britain

Date of deposit	Accession Number
29-Feb-96	ECACC 96022936

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

"In respect of all designated states in which such action is possible and to the extent that it is legally permissible under the law of the designated state, it is requested that a sample of the deposited micro-organism(s) be made available only by the issue thereof to an independent expert, in accordance with the relevant patent legislation, *e.g.* EPC Rule 28(4), U.K. Rule 17(3), Australian Regulation 3.25(3) and generally similar provisions *mutatis mutandis* for any other designated state".

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 17, line 1-4

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

National Collection of Industrial and Marine Bacteria

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen
Scotland AB2 1RY

Date of deposit	Accession Number
23-Nov-94	NCIMB 40694

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

"In respect of all designated states in which such action is possible and to the extent that it is legally permissible under the law of the designated state, it is requested that a sample of the deposited micro-organism(s) be made available only by the issue thereof to an independent expert, in accordance with the relevant patent legislation, *e.g.* EPC Rule 28(4), U.K. Rule 17(3), Australian Regulation 3.25(3) and generally similar provisions *mutatis mutandis* for any other designated state".

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 29, line 14-16

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

National Collection of Industrial and Marine Bacteria

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen
Scotland AB2 1RY

Date of deposit	Accession Number
11-Apr-96	NCIMB 40799

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

"In respect of all designated states in which such action is possible and to the extent that it is legally permissible under the law of the designated state, it is requested that a sample of the deposited micro-organism(s) be made available only by the issue thereof to an independent expert, in accordance with the relevant patent legislation, *e.g.* EPC Rule 28(4), U.K. Rule 17(3), Australian Regulation 3.25(3) and generally similar provisions *mutatis mutandis* for any other designated state".

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

CLAIMS

1. A modified prodomain of carboxypeptidase B which enhances recombinant expression thereof when co-expressed from a separate gene.

2. A modified prodomain according to claim 1 wherein the prodomain is modified at its C-terminus.

3. A modified prodomain according to claim 2 wherein the C-terminus is modified by addition of at least one amino acid.

4. A modified prodomain according to claim 3 wherein the C-terminus is modified by addition of 1-20 amino acid(s).

5. A modified prodomain according to claim 3 wherein the C-terminus is modified by addition of 1-6 amino acid(s).

6. A modified prodomain according to any one of claims 3-5 wherein the C-terminus amino acid of the prodomain after addition of the amino acid(s) is a hydrophobic amino acid.

7. A modified prodomain according to claim 6 wherein the hydrophobic amino acid is selected from leucine, isoleucine, valine, alanine or phenylalanine.

8. A modified prodomain according to claim 5 wherein the added amino acid(s) are any one of the following: L, KDEL, KKAA or SDYQRL.

9. A modified prodomain of carboxypeptidase B according to any preceding claim in which the carboxypeptidase is human pancreatic carboxypeptidase B.

10. A modified prodomain according to claim 8 which is a human pancreatic carboxypeptidase B prodomain with a leucine added at its C-terminus.

11. A polynucleotide sequence capable of encoding a modified prodomain as defined in any one of claims 1-10.

12 A vector comprising a polynucleotide sequence as defined in claim 11.

13. A host cell comprising a polynucleotide sequence as defined in claim 11.

14. A method of recombinant carboxypeptidase B production which comprises simultaneously expressing in a eucaryotic host cell a carboxypeptidase B enzyme gene together with a separate gene encoding a modified prodomain as defined in any one of claims 1-10 and optionally at least partially purifying the recombinant carboxypeptidase B.

15. A method according to claim 14 in which the eucaryotic host cell is mammalian and:

i) the recombinant carboxypeptidase B is in the form of a humanised 806.077 $F(ab')_2$ - $\{[A248S,G251T,D253K]HCPB\}_2$ fusion protein;

ii) the carboxypeptidase B enzyme gene is in the form of a gene encoding a humanised Fd heavy chain fragment of antibody 806.077 linked to enzyme [A248S,G251T,D253K]HCPB;

iii) the separate gene encoding a modified prodomain encodes human pancreatic carboxypeptidase B prodomain with a leucine added at its C-terminus; and

iv) a further gene is co-expressed which encodes a humanised light chain of antibody 806.077;

and wherein the fusion protein is in the form of a $F(ab')_2$ with a molecule of [A248S,G251T,D253K]HCPB at a C-terminus of each of its heavy chain fragments.

Figure 1

Human Pancreatic Carboxypeptidase B

Cloning

Pancreas carboxypeptidase B.

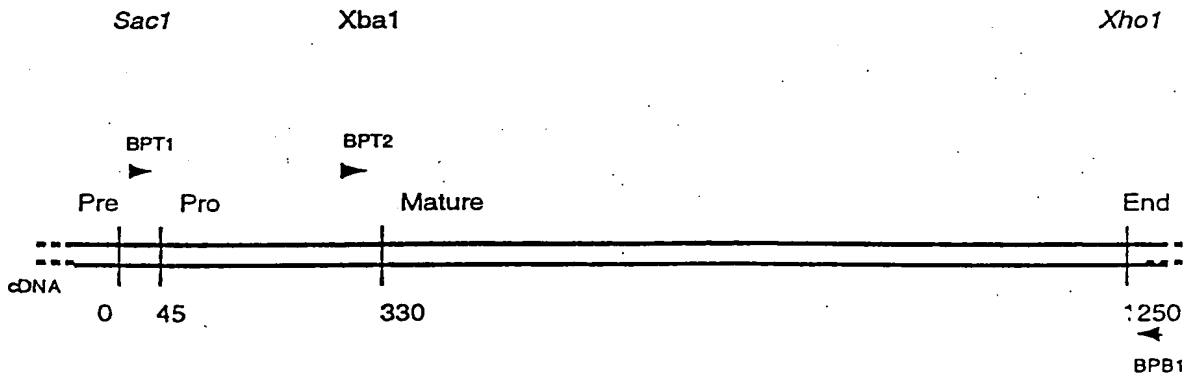
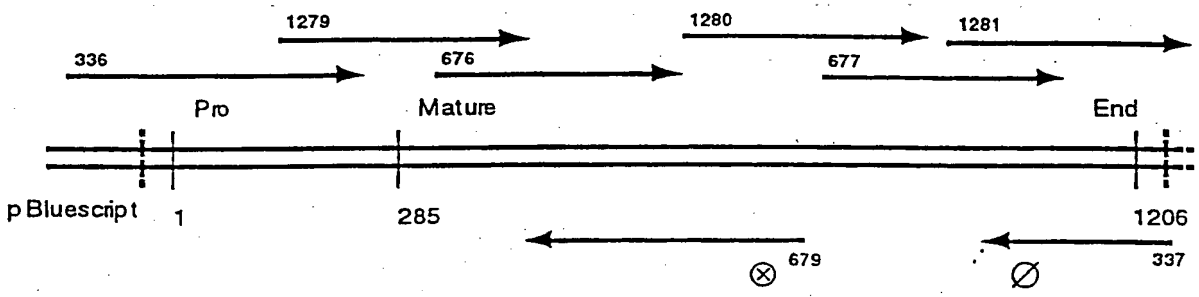


Figure 2

Human Pancreatic Carboxypeptidase B

Sequencing



All 6 clones have identical sequence, and all have :-

∅ Aspartate in the enzyme recognition site. ie Carboxypeptidase B.

When compared with published sequence :-

⊗ TGC codon insert, changing polypeptide ...GSSIG... to ...GSSCIG... .

Figure 3

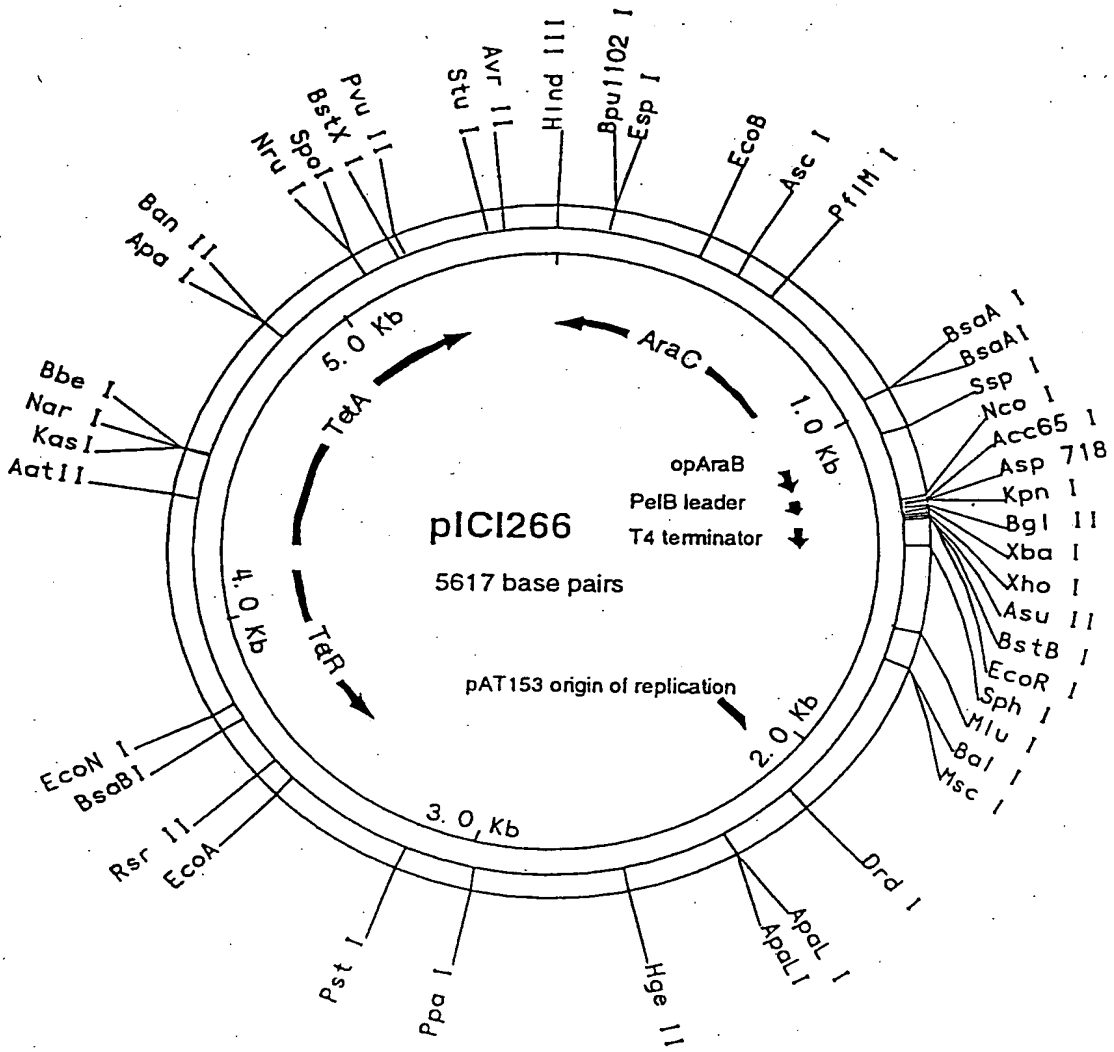
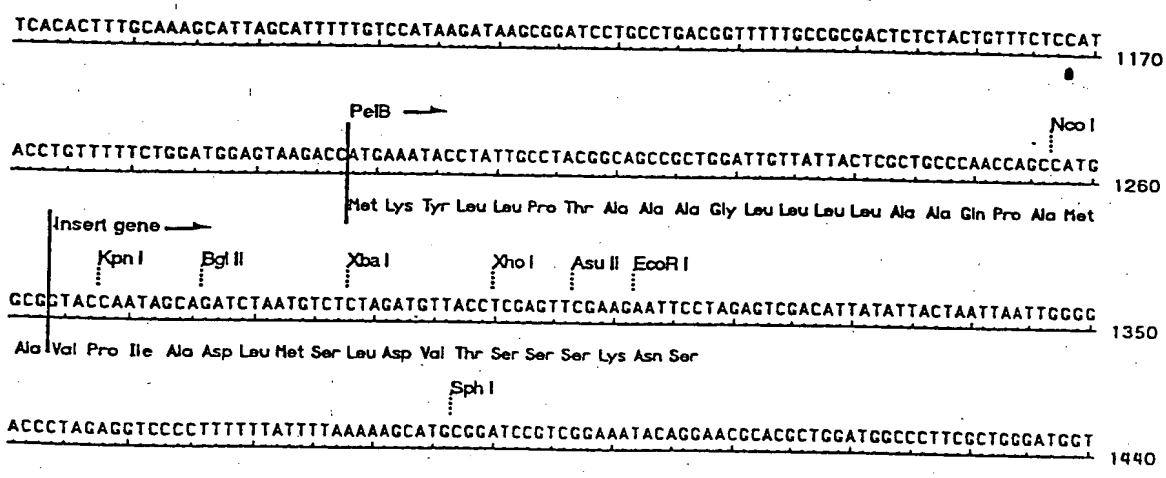


Figure 4

pICI266 expression vector - gene cloning



INTERNATIONAL SEARCH REPORT

In. .ational Application No
PCT/GB 98/00415

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C07K14/47 C12N15/57 C12N15/63

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 97 42329 A (ZENECA LIMITED) 13 November 1997 *see the whole patent*	1-15
X	WO 96 20011 A (ZENECA LIMITED) 4 July 1996 *see the whole patent*	1-15
A	EP 0 585 570 A (BEHRINGWERKE) 9 March 1994 *see the whole patent*	15

[ ] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 June 1998	09/07/1998

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer
Marie, A

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. ational Application No
PCT/GB 98/00415

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9742329 A	13-11-1997	AU 2645597 A	26-11-1997
WO 9620011 A	04-07-1996	AU 4269796 A	19-07-1996
		CA 2205091 A	04-07-1996
		CZ 9701952 A	17-12-1997
		EP 0806964 A	19-11-1997
		FI 972683 A	19-06-1997
		GB 2309033 A	16-07-1997
		HU 77450 A	28-04-1998
		NO 972882 A	19-08-1997
		PL 320964 A	24-11-1997
		SK 80997 A	04-02-1998
		ZA 9510918 A	24-06-1996
		AU 6747696 A	19-03-1997
		EP 0844885 A	03-06-1998
		WO 9707769 A	06-03-1997
		GB 2317387 A	25-03-1998
		NO 980630 A	07-04-1998
EP 585570 A	09-03-1994	DE 4225853 A	10-02-1994
		AU 4441393 A	10-02-1994
		CA 2101868 A	06-02-1994
		JP 6205693 A	26-07-1994
		US 5645817 A	08-07-1997